

# Civil Rights Perestroika: Intergroup Relations after Affirmative Action

Linda Hamilton Krieger†

## TABLE OF CONTENTS

Introduction .....	1254
I. Preferences and Social Perception .....	1258
A. Preferences and Social Cognition: The Problems of Stigma and Self-Derogation.....	1259
1. Self-Evaluations of Aptitude and Achievement.....	1261
2. The Effect of Preferences on Motivation, Commitment, and Task Selection.....	1262
B. Preferences and Social Perception: Evaluation, Attribution, and the Perpetuation of Negative Stereotypes .....	1263
1. Preferential Selection and the Evaluation of Affirmative Action Beneficiaries.....	1264
2. Affirmative Action Preferences and the Perpetuation of Negative Stereotypes: The Role of Attributional Ambiguity.....	1265
C. Preferences and Procedural Justice.....	1271
D. Preferences and Social Identity.....	1274
II. After Affirmative Action: Recognizing Discrimination in the Land of the Colorblind.....	1276
A. The Inefficacy of Colorblindness as a Normative Construct.....	1277
1. Automatic Processes in Intergroup Judgment.....	1279
2. Controlled Processes and Nondiscrimination: Taming the Beast of Automaticity .....	1285
a. Unawareness of Mental Process.....	1287
b. The Role of Motivation .....	1289

---

Copyright © 1998 California Law Review, Inc.

† Acting Professor, University of California School of Law (Boalt Hall); A.B. Stanford University, 1975; J.D. New York University School of Law, 1978. This project has profited from the assistance of many people. Special thanks are owed to Jan Vetter, Stephen McGeorge Bundy, James S. Uleman, Sheldon Zedeck, and the members of the Boalt Hall Junior Faculty Workshop.

c. The Limits of Mental Correction .....	1290
B. Merit.....	1291
1. The Merit Principle and Perceptions of Procedural Justice.....	1291
2. Social Identity, Social Cognition, and the Construction of Merit.....	1293
3. Intergroup Bias in the Construction of Merit: The Inadequacy of Existing Law .....	1298
C. The Limits of Individualized Adjudication.....	1302
1. Individualized Adjudication and the Problem of Case- By-Case Data.....	1305
2. Individual Adjudication and the Problem of Schematic Expectancies .....	1309
a. Schematic Expectancies and the Problem of Complexity .....	1312
b. Schematic Expectancies and the Problem of Ingroup Favoritism.....	1317
c. In-Group Favoritism and the Limitations of Current Law .....	1325
D. Individualized Adjudication, Hypothesis Testing, and Causal Attribution: The Effects of Intergroup Bias .....	1327
Conclusion: Social Cognition, Social Identity, and the Ecology of Intergroup Bias.....	1329

## Civil Rights Perestroika: Intergroup Relations after Affirmative Action

Linda Hamilton Krieger

Change? Yes, we must change, only show me the Theory, and I will be at the barricades, show me the book of the next Beautiful Theory, and I promise you these blind eyes will see again, just to read it, to devour that text. Show me the words that will reorder the world, or else keep silent.<sup>1</sup>

*In this Article, Professor Krieger examines the implications of social cognition and social identity theory for the debate over affirmative action. In Part I, she explores the extent to which insights from those fields support the claim that affirmative action preferences exacerbate intergroup tensions and perpetuate certain subtle forms of intergroup bias. Finding qualified support for that view in both theoretical models and empirical evidence, Part I concludes that at least certain preferential forms of affirmative action may injure intergroup relations in a variety of troubling ways.*

*Extending the analysis in Part II, Professor Krieger inquires whether, absent preferential forms of affirmative action, remaining legal and policy tools will suffice to control discrimination and prevent the further segregation of American society. Part II concludes that these remaining tools, which include a colorblindness model of nondiscrimination, reliance on an objective concept of merit, and the use of individual disparate treatment adjudication as a primary law enforcement tool, are unequal to the task. The misplaced confidence in these tools often found among affirmative action opponents derives, Krieger suggests, from a misunderstanding of the nature and sources of intergroup bias, from a failure to recognize its subtlety and tendency to persist over time, and from over-reliance on limited adjudicatory and regulatory*

---

1. TONY KUSHNER, *ANGELS IN AMERICA, PART II: PERESTROIKA*, Act I, Scene 1 (1994). *Perestroika* (restructuring) was the term attached to Mikhail Gorbachev's efforts during the latter half of the 1980s to dismantle the Soviet Union's command and control economy. The policy was frequently criticized on the grounds that dismantling state controls before the institutions, norms, and other structures required for a free market system had been developed would leave the Union vulnerable to political, economic, and ultimately social chaos.

*approaches to address what is fundamentally a complex cultural problem.*

*Accordingly, Professor Krieger argues that we are not yet ready to abandon preferential forms of affirmative action for the simple reason that we have nothing adequate with which to replace them. Unless more inclusive jurisprudential models of intergroup bias and new approaches to reducing such bias are developed, the problems of discrimination and inequality of opportunity can be expected to worsen in a post-affirmative action environment. In her Conclusion, Krieger articulates a set of first principles and constructs a general conceptual foundation for the future development of such a broadened view.*

### INTRODUCTION

Here at the University of California at Berkeley, there was a surreal quality to November 6, 1996, the day after voters, in enacting Proposition 209, elected to end affirmative action in California state hiring, contracting, and education. The few protests that had been organized ended quickly and quietly. At the law school, students seemed uncharacteristically subdued: quietly resigned—or quietly pleased. In the newspapers that morning and on the mornings that followed, articles about Texaco executives referring to their African-American employees as “niggers”<sup>2</sup> and “black jelly beans” were oddly juxtaposed against others in which Proposition 209’s triumphant sponsors heralded a glorious new era of truly equal opportunity and the long-awaited dawning of a colorblind society.

Now, many months later and well into the first of Berkeley’s “post-affirmative action” years, the atmosphere here is no less strange. On the first day of classes in 1997, the halls and courtyards swarmed with television cameras and reporters, attempting, I noted with a sense of irony, to identify the lone African-American member of Boalt Hall’s first “colorblind” class. We were no longer supposed to consider race, but race was everywhere in these halls on that first morning of the new school year.

The weeks that followed provided little respite. Three days after classes commenced, the Ninth Circuit reaffirmed a three judge panel’s

---

2. I refer here not to those portions of certain tape recorded conversations that were eventually deciphered as making reference to “Saint Nicholas,” but rather to the comments of Texaco Assistant Controller Jim Wooley, as described in deposition testimony of Texaco manager Michael Moccio and reported in the *New York Times*. When asked by Mr. Moccio what he should do about a race discrimination claim filed by an African-American employee, Mr. Wooley reportedly replied, “I’d fire her black ass.” When Mr. Moccio protested that firing the employee would constitute unlawful retaliation, Mr. Wooley rejoined, “I guess we treat niggers differently down here.” Kurt Eichenwald, *The Two Faces of Texaco*, N.Y. TIMES, Nov. 10, 1996, at Section 3, Page 1.

prior decision upholding Proposition 209<sup>3</sup> and refused to consider its constitutionality en banc.<sup>4</sup> The following Tuesday, the same court denied plaintiff's motion for an emergency stay pending petition to the Supreme Court, and Proposition 209 became law.<sup>5</sup> Eight days later, the same Ninth Circuit panel that originally upheld 209 ruled that the Equal Protection Clause prohibits not only the use of overt preferences, but also requirements that government contractors make "good faith efforts" to recruit women and minority subcontractors.<sup>6</sup> On November 4, 1997, the Supreme Court denied certiorari in *Coalition for Economic Equity v. Wilson*.<sup>7</sup> The legal fight over Proposition 209 was over.

If history is any guide, the trend started in California will spread to other states and to the national stage in the months ahead. The 105<sup>th</sup> Congress witnessed the introduction of three separate bills that would have "nationalized" Proposition 209,<sup>8</sup> and, events in Houston notwithstanding,<sup>9</sup> many States are contemplating similar legislation.<sup>10</sup>

But before proceeding further down this road, it might be wise to pause and ask some hard questions. As a society, are we really ready to abandon preferential forms of affirmative action?<sup>11</sup> What can we expect

3. See *Coalition for Economic Equity v. Wilson*, 110 F.3d 1431 (9th Cir. 1997).

4. See *Coalition for Economic Equity v. Wilson*, 122 F.3d 692 (9th Cir. 1997).

5. See *Coalition for Economic Equity v. Wilson*, 122 F.3d 718 (9th Cir. 1997).

6. See *Monterey Mechanical Co. v. Wilson*, 125 F.3d 702 (9th Cir. 1997).

7. See 118 S. Ct. 397 (1997).

8. See Civil Rights Restoration Act of 1997, S. 46 105th Cong. (1997); Civil Rights Act of 1997, H.R. 1909, 105th Cong. (1997); Racial and Gender Preference Reform Act, H.R. 2079, 105th Cong. (1997).

9. On November 4, 1997, voters in Houston, Texas defeated Measure A, which would have prohibited affirmative action in city hiring and contracting. See Sam Howe Verhovek, *Voters Turn Back Affirmative Action Ban Much Like the One in California*, N.Y. TIMES, Nov. 5, 1997, at A24.

10. At last count, at least 26 states were contemplating similar legislation. See Harriet Chiang, *U.S. High Court Allows Prop. 209 to Remain in Effect*, S.F. CHRON., Sept. 5, 1997, at A17. Among the most successful so far is Initiative 200 in Washington, dubbed the "Civil Rights Initiative" by supporters. The initiative, which was certified in January after supporters submitted more than 280,000 signatures, goes next to the state legislature and then on to the voters of Washington. See David Postnan, *Enough Valid Signatures Certified for Initiative 200: Proposal to End Affirmative Action in Government Goes to Legislature*, SEATTLE TIMES, Jan. 21, 1998, at B2. Colorado is expected to have an initiative similar to Proposition 209 on its November 1998 ballot. See Beverly Vasquez, *Colorado Unity Defends Affirmative Action*, DENVER BUS. J., Nov. 28, 1997, at 15A. In Georgia and Florida, opponents of affirmative action have not enjoyed similar success. In Georgia the leading proposal against affirmative action has faltered in a committee of the state House since 1995. See David Pendered, '98 Georgia Legislature: Anti-Affirmative Action Advocates See '98 as Theirs, THE ATLANTA CONSTIT., Jan. 15, 1998, at C4. Supporters of the Florida initiative modeled after Prop. 209 have collected only one percent of the votes needed to place it on the ballot. See Gady A. Epstein, *Anti-Affirmative Action Petition Founders*, TAMPA TRIB., Dec. 30, 1997, at Metro 1.

11. Affirmative action does not necessarily involve the preferential use of race, gender, or ethnicity in selection decision making. As originally defined, affirmative action simply requires organizations to ensure that their selection procedures are free of subtle forms of discrimination and that they take "extra" or "affirmative" steps to increase the representation of previously under-represented groups in their applicant pools. See U.S. Commission on Civil Rights, *Statement on Affirmative Action* (Report No. 54, 1977). For ease of expression, I will at times use the phrase

to occur with respect to intergroup relations, and with race relations in particular, in a post-affirmative action environment? If affirmative action is eliminated, will remaining policy tools prove adequate to effectuate racial and gender equity and to prevent the resegregation of American society? Will the idea of "colorblindness" suffice as a theoretical model for understanding what it means not to discriminate? Or will we find instead that affirmative action actually served to mask a "multiplicity of sins"—critical failings in our approach to intergroup relations and serious defects in the tools available to make the equal opportunity society a reality instead of a hazy, unattainable dream?

In the face of voter response, supporters of affirmative action too face hard questions. If, as social scientists have shown, people's perceptions of fairness influence their attitudes and behaviors, and if, as the vote on Proposition 209 reflects, large segments of the population view racial preferences as unfair, is there not substantial reason to fear that preferential forms of affirmative action might ultimately exacerbate not only racial tensions, but also racial bias itself? However well-intentioned, might preferences heighten intergroup tensions, reinforce stereotyped notions of outgroup inferiority, and ultimately undermine the moral authority of equal opportunity law?

Many have written about affirmative action. The issue can be approached from a variety of disciplinary perspectives: legal, philosophical, or political-theoretical, to name but a few. But whether supportive or critical, and irrespective of the discipline from which they derive, most treatments either explicitly advance or implicitly invoke a set of essentially empirical claims. Some of these claims concern the effect of affirmative action programs on beneficiaries or on patterns of intergroup relations more generally. For example, affirmative action opponents commonly argue that group preferences actually injure beneficiaries, both by creating self-doubt and by provoking negative responses in non-beneficiaries. Supporters of affirmative action, on the other hand, dispute these claims, attacking the efficacy of an equal treatment approach to equal opportunity and intergroup justice.

In this Article, I do not intend to evaluate the constitutional or statutory validity of affirmative action preferences, nor to respond to philosophical or political-theoretical arguments supporting or attacking them. Rather, my effort here centers on a close examination of the essentially empirical psychological claims around which much of the debate revolves. This project strikes me as important for a number of reasons.

---

"affirmative action" to refer to the overt, preferential use of group membership in selection decision making. However, the reader should understand that affirmative action and the use of group-based preferences are not coextensive.

If we accept the premise that at least one important goal of civil rights policy is to lessen the tendency toward intergroup strife inherent in any diverse society, then one can assess the value of a particular policy tool only in light of its expected effect on people's attitudes and behaviors towards members of other social groups. Similarly, if we are to eliminate preferential forms of affirmative action, we would be well-advised to consider whether remaining legal policy tools, such as the "colorblindness" approach to nondiscrimination, will adequately control intergroup bias and ensure equality of opportunity for members of all gender, ethnic, and racial groups. To make such assessments, one must know something about the psychology of social perception and judgment generally, and more particularly about social perception and judgment in intergroup contexts. A civil rights policy uninformed by relevant insights from social psychology may fall prey to the law of unintended consequences.

In this Article, I will evaluate preferential forms of affirmative action, and the "colorblindness" approach to nondiscrimination championed by its opponents, through the lens of three sub-disciplines of social psychology. The first, social cognition theory, concerns how patterns of human cognition, while adaptive in many important respects, lead to systematic errors which bias social perception and judgment in a variety of predictable ways. The second, social justice theory, examines the determinants and consequences of people's subjective judgments of fairness. The third, social identity theory, seeks to describe and explain how and why people tend to favor members of their own groups—even when membership in those groups is based on trivial or arbitrary criteria.

My exploration divides into two broad parts. Part I explores the extent to which insights from cognitive social psychology reasonably support arguments that affirmative action preferences exacerbate intergroup tensions and perpetuate certain subtle forms of intergroup bias.

Such arguments are based on a variety of essentially empirical claims, including: (a) that preferences lead to self-derogation and disaffection among affirmative action beneficiaries; (b) that preferences perpetuate negative stereotypes about beneficiary groups and retard the process of stereotype change; (c) that preferences, by virtue of their perceived unfairness, create intergroup hostility; and (d) that the use of race, gender, or national origin in selection decision making makes those categories more rather than less salient in social perception and judgment.

Most of these hypotheses about affirmative action's negative effects can be tested, and a good deal of empirical data exists to support or refute them. After reviewing the relevant literature, and exploring

both its implications and its limitations, Part I concludes that certain preferential forms of affirmative action can be expected, at least under some conditions, to injure intergroup relations in a variety of troubling ways.

Part II inquires whether, absent preferential forms of affirmative action, remaining policy tools would prove adequate to control discrimination and prevent the further segregation of American society. It concludes that these remaining tools, which include a colorblindness model of nondiscrimination, an objective concept of merit, and individualized adjudication as a primary policy enforcement tool, are unequal to the task. In short, we still lack adequate tools for coping successfully with the problems of intergroup competition and cooperation in a pluralistic society.

This failure, which I will argue has been masked to a certain extent by preferential forms of affirmative action, derives from a misunderstanding of the nature and sources of intergroup bias, from a failure to recognize its tendency to persist over time, and from over-reliance on limited adjudicatory and regulatory approaches to address what is fundamentally a complex cultural problem. Accordingly, I argue that unless we develop a broadened understanding of intergroup bias and new approaches to reducing it, the problems of discrimination and inequality of opportunity will worsen in a post-affirmative action environment.

In this Article's Conclusion, I seek to construct a general conceptual foundation for the future development of this broadened view. Its central premise is simple: discrimination does not solely derive from stable, dispositional traits internal to actors we call "discriminators." Rather, intergroup bias increases or decreases in response to contextual, environmental factors which shape how social actors perceive, judge, and make decisions about members of their own and other social reference groups. Accordingly, an anti-discrimination policy grounded in an individualized search for discriminatory intent cannot be expected to succeed either in identifying and preventing intergroup bias or in managing social tendencies toward intergroup conflict. If we are to solve the problem of intergroup discrimination, we must attend more closely to the ecology of intergroup relations. Eliminating affirmative action before we have developed an effective alternative theoretical and doctrinal approach to managing intergroup bias is a strategy more risky than many might assume.

## I

### PREFERENCES AND SOCIAL PERCEPTION

Instrumental arguments against affirmative action are based at least in part on the essentially empirical claim that ultimately, preferences



do more harm than good. Such arguments divide into three broad categories: arguments about stigma, arguments about the perpetuation of negative stereotypes and other subtle forms of bias, and arguments about the negative effects of gender or race-based classifications on intergroup relations more generally. While most policy-oriented social cognition literature supports affirmative action, one finds in both the theoretical models of intergroup behavior and the limited empirical evidence that exists, reasonable grounds for concern that at least in some circumstances preferences may have detrimental effects on intergroup relations.

A. *Preferences and Social Cognition:  
The Problems of Stigma and Self-Derogation*

Among affirmative action critics, one finds widespread adherence to the idea that using racial or gender preferences stigmatizes beneficiaries and ultimately undermines their self-confidence and self-esteem.<sup>12</sup> In at least one version of this argument, critics maintain that giving a preference to members of a particular social group necessarily compromises the principle of merit-based decision making. To the extent that non-merit criteria are used, selected members of preferred groups will be unsure about the reasons for their success. This uncertainty creates a sense of self-doubt and self-derogation in affirmative action beneficiaries.<sup>13</sup>

Although various high-profile affirmative action beneficiaries have compellingly and passionately advanced this claim,<sup>14</sup> research findings supporting it are complex, contradictory, and difficult to interpret. Complicating matters further is the unfortunate fact that most research on the self-derogating effects of affirmative action has focused on gender preference. Little empirical work has been done on the effects of preferential selection on minority self-evaluation.

The evidence that does exist suggests that, at least under certain conditions, preferential selection may negatively affect women's self-evaluations, commitment, and task selection. Thus in a series of studies, Madeline Heilman and her colleagues demonstrated that women led to

---

12. See, e.g., Midge Decter, *Benign Victimization*, 13 POLICY REVIEW 65 (1980); Jeff Howard & Ray Hammond, *Rumors of Inferiority: The Hidden Obstacles to Black Success*, NEW REPUBLIC 17 (Sept. 1985); Leonard Reed, *What's Wrong with Affirmative Action*, WASH. MONTHLY 24 (Jan. 1981); Thomas Sowell, *The Plight of Black Students in America*, 103 DAEDALUS 179 (1974).

13. The classic statement of this argument can be found in SHELBY STEELE, *THE CONTENT OF OUR CHARACTER* (1990). Steele writes, "The effects of preferential treatment—the lowering of normal standards to increase black representation—puts blacks at war with an expanded realm of debilitating doubt, so that the doubt itself becomes an unrecognized preoccupation that undermines their ability to perform." *Id.* at 117.

14. See, e.g., STEPHEN L. CARTER, *REFLECTIONS OF AN AFFIRMATIVE ACTION BABY* (1991); STEELE, *supra* note 13. See also sources cited *supra*, note 12.

believe that they had been selected for a particular position because of their gender subsequently evaluated their skills in the relevant domain more harshly,<sup>15</sup> showed less interest in persisting in the position,<sup>16</sup> chose less challenging tasks,<sup>17</sup> and evaluated other female but not male applicants more negatively than did female subjects who believed they had been selected using merit-based criteria.<sup>18</sup>

Other researchers have observed similar effects. For example, a 1989 study asked male and female subjects to evaluate their own performance on a creativity task. Researchers told one set of subjects that "affirmative action" had been used to insure that enough members of their gender would be included in the study. Researchers told other subjects that they had been included because they had received a high score on a previously-administered creativity test. Although there were no significant effects on measured creativity, subjects in the "affirmative action" condition evaluated their performance significantly less favorably than did subjects in the "merit" condition.<sup>19</sup>

In an earlier study, William Austin and his colleagues assessed the reactions of male and female beneficiaries in a simulated affirmative action selection procedure.<sup>20</sup> In the "affirmative action" condition, subjects were told that because an insufficient number of members of the subjects' gender group had previously been selected to receive a desired benefit, they would now be selected to receive it. In a second condition, subjects were told that their receipt of the benefit was based on performance.

In Austin's original study, subjects of both genders selected in the "affirmative action" condition evaluated their selection as unfair and reported moderate levels of emotional discomfort and distress. Interestingly, however, subsequent attempts at replication found white males apparently immune to the effect. In these studies, only women and

---

15. See Madeline E. Heilman et al., *Intentionally Favored, Unintentionally Harmed? Impact of Sex-Based Preferential Selection on Self-Perceptions and Self-Evaluations*, 72 J. APPLIED PSYCHOL. 62 (1987). See also Marlene E. Turner et al., *Sex Differences in Reactions to Preferential Selection: Towards a Model of Preferential Selection as Help*, 6 J. SOC. BEHAV. & PERSONALITY 797 (1991) (replicating Heilman's results).

16. See Heilman et al., *supra* note 15.

17. See Madeline E. Heilman et al., *Skirting the Competence Issue: Effects of Sex-Based Preferential Selection on Task Choices of Women and Men*, 76 J. APPLIED PSYCHOL. 99 (1991).

18. See Madeline E. Heilman, Jonathan A. Lucas, & Stella R. Kaplow, *When Similarity Is a Liability: The Effects of Sex-Based Preferential Selection on Reactions to Like-Sex and Different-Sex Others*, 78 J. APPLIED PSYCHOL. 917 (1991).

19. See Rupert W. Nacoste, *Affirmative Action and Self-Evaluation*, in AFFIRMATIVE ACTION IN PERSPECTIVE 103 (Fletcher A. Blanchard & Faye J. Crosby, eds., 1989).

20. See William Austin et al., *Responses to Favorable Sex Discrimination*, 1 LAW & HUM. BEHAV. 283 (1977).

members of ethnic groups operating under a "suspicion of inferiority"<sup>21</sup> reported feeling stigmatized by preferential selection.<sup>22</sup>

One can certainly find support in these studies for the argument that preferences injure beneficiaries by creating, or by reinforcing, an internal sense of stigma and self-doubt. But the picture is not so simple. Subsequent studies have demonstrated that the self-denigrating effects of affirmative action are highly sensitive to contextual variables and, under certain conditions, disappear entirely. To appreciate the complexity of the empirical findings on this issue, it is useful to divide the research into two broad areas: the effects of preferences on self-evaluations of aptitude and achievement; and the effects of preferences on motivation, commitment, and task selection.

### *1. Self-Evaluations of Aptitude and Achievement*

Subsequent research has significantly qualified early research results indicating that preferential selection leads to a self-demigrating evaluation effect. For example, where beneficiaries receive positive, unambiguous performance feedback, the self-denigrating effect disappears—at least as to those aspects of performance on which feedback was received.<sup>23</sup> But this feedback generalizes neither to other aspects of performance, nor to broader self-conceptions of ability.<sup>24</sup>

To complicate matters further, some studies indicate that the effects of preferential selection on self-evaluation vary according to the precise nature of the selection procedure. Thus, in studies using gender alone to make selections, there is a relatively strong negative self-evaluation effect.<sup>25</sup> But the effect seems to disappear when subjects are

---

21. See Brenda Major et al., *Attributional Ambiguity of Affirmative Action*, 15 BASIC & APPLIED SOC. PSYCHOL. 113, 118 (1994).

22. See Heilman et al., *supra* note 15, at 67 (reporting that women's, but not men's, self-perceptions and self-evaluations were negatively affected by a sex-based preferential selection method relative to a merit-based selection method); Madeline E. Heilman et al., *Self-Derogating Consequences of Sex-Based Preferential Selection: The Moderating Role of Initial Self-confidence*, 46 ORGANIZATIONAL BEHAV. & HUM. DECISION PROCESSES 202 (1990) (reporting that when provided with no positive or negative information about performance on a test for leadership, female subjects selected preferentially reported lower self-evaluations than females selected under a merit condition; no such differences in self-evaluation were reported observed in male subjects across the preferential and merit selection conditions); Monica E. Schneider et al., *Social Stigma and the Potential Costs of Assumptive Help*, 22 PERSONALITY & SOC. PSYCHOL. J. 201, 207 (1996) (reporting that black subjects receiving unrequested help from a white peer had lower competence-based self esteem and more depressed affect than white subjects who received unrequested help; no differences in self-esteem or affect were found between blacks and whites who did not receive help).

23. See Turner et al., *supra* note 15, at 810.

24. See *id.* For further discussion of these findings, see Marlene E. Turner & Anthony R. Pratkanis, *Affirmative Action as Help: A Review of Recipient Reactions to Preferential Selection and Affirmative Action*, 15 BASIC & APPLIED SOC. PSYCHOL. 43, 47 (1994).

25. See sources cited *supra* note 15.

told that qualifications, as well as group membership, have been used in making selections.<sup>26</sup> Compounding the observed complexity, other investigators find that selection procedure has no effect on patterns of self-evaluation in affirmative action beneficiaries,<sup>27</sup> or find that subjects' personal views about the fairness of affirmative action determine whether they experience a self-denigrating effect of preferential selection.<sup>28</sup> In short, evidence on this question is mixed and the findings difficult to interpret.

## 2. *The Effect of Preferences on Motivation, Commitment, and Task Selection.*

Subsequent research has similarly qualified empirical findings indicating that preferential selection negatively affects beneficiaries' motivation, commitment, and task selection. For example, individual qualifications along with group membership criteria lessened the negative effect of preferential affirmative action on women's interest in affiliating themselves with a particular employer.<sup>29</sup> Similarly, female subjects' interest in entering an occupation in which gender preferences are used in hiring varies with the proportional representation of women in that occupation.<sup>30</sup> In another study, while women selected on sex alone chose less demanding performance tasks, women selected on the basis of both sex and qualifications demonstrated no such effect.<sup>31</sup> Finally, while one study found an affirmative action program corresponded with lower levels of job excitement among black males, it found precisely the opposite effect for black females.<sup>32</sup> Thus, studies differ in their findings on the effects of preference on beneficiary motivation, commitment, and task selection,<sup>33</sup> with results depending largely on contextual variables.<sup>34</sup>

---

26. See Heilman et al., *supra* note 15; Major et al., *supra* note 21; Rupert W. Nacoste, *Selection Procedure and Responses to Affirmative Action*, 9 LAW & HUM. BEHAV. 225 (1985).

27. See Turner et al., *supra* note 15.

28. See Nacoste, *supra* note 19, at 108.

29. See Nacoste, *supra* note 26.

30. See Madeline E. Heilman & Joyce Mardenfeld Herlihy, *Affirmative Action, Negative Reaction? Some Moderating Conditions*, 33 ORGANIZATIONAL BEHAV. & HUM. PERF. 204, 208 (1984).

31. See Heilman et al., *supra* note 17.

32. See Marylee C. Taylor, *Impact of Affirmative Action on Beneficiary Groups: Evidence from the 1990 General Social Survey*, 15 BASIC & APPLIED SOC. PSYCHOL. 143 (1994).

33. See Heilman et al., *supra* note 15; Heilman et al., *Self-Derogating Consequences*, *supra* note 22; Heilman et al., *Skirting the Competence Issue*, *supra* note 17.

34. See Turner et al., *supra* note 15; Marlene E. Turner & Anthony R. Pratkanis, *Effects of Preferential and Meritorious Selection on Performance: An Examination of Intuitive and Self-Handicapping Perspectives*, 19 PERSONALITY & SOC. PSYCHOL. BULL. 47 (1993) (including variables such as sex-based selection and task motivation).

In short, empirical investigations of the self-derogating effects of affirmative action so passionately described in the popular literature paint no clear picture and suggest no clear policy direction. On the one hand, some studies indicate a reasonable basis for concern. On the other hand, subsequent findings suggest that beneficiaries' responses will vary considerably depending on who the beneficiaries are, how they feel about affirmative action, and how specific policies are described and implemented. Research findings so complex and contradictory provide a poor foundation for public policy making, and from the standpoint of advocates on either side of the debate, can probably best be described as unhelpful.

*B. Preferences and Social Perception: Evaluation, Attribution, and the Perpetuation of Negative Stereotypes*

Affirmative action opponents frequently argue that using racial or gender preferences reinforces negative stereotypes about minorities and women. The argument generally proceeds as follows: nonbeneficiaries who suspect that preferences played a part in decision making will assume that merit-based criteria alone would not have selected preferred group members. Because preference permits people to attribute preferred group members' success to factors other than merit, affirmative action reinforces stereotypic assumptions of outgroup inferiority and thus hampers stereotype change.

In evaluating the utility of affirmative action preferences as a tool for lessening intergroup strife, it would be valuable to know whether preferences do reinforce negative stereotypes. Whether we think the use of preferences *should* reinforce negative stereotypes of women and minorities, it is important to know whether they *do*.

Research findings on this question are somewhat more straightforward and provide a more substantial basis for concern than those relating to affirmative action's potential self-derogating effects. Indeed, the weight of empirical evidence reasonably supports the view that using preferences tends to exacerbate subtle forms of intergroup bias in the evaluation of affirmative action beneficiaries. Theoretical interpretations also provide substantial basis for concern: much of what we know about stereotypes' tendency to resist the corrective influence of disconfirming evidence reinforces concerns about the negative effects of affirmative action on intergroup perception and judgment.

To understand these empirical findings and explanatory models, and to appreciate their significance for the affirmative action debate, it is useful to consider research on the effects of preferences in three

broad areas: the evaluation of beneficiary qualifications, causal attribution,<sup>35</sup> and stereotyping and stereotype change.

### *1. Preferential Selection and the Evaluation of Affirmative Action Beneficiaries*

A substantial body of research considers whether preferences negatively affect other people's evaluations of affirmative action beneficiaries. In one group of studies, experimenters asked subjects to evaluate the qualifications of people supposedly selected for employment or for admission to programs of higher education. Researchers indirectly informed subjects in one condition that the selecting institution had an affirmative action program. In a second condition, researchers made no mention of affirmative action. Subjects in the "affirmative action" condition consistently rated the files of selected women and minorities as reflecting lower levels of competence, qualification, and accomplishment than did subjects evaluating identical files in the "non-affirmative action" condition.<sup>36</sup>

In a related study, subjects paired five resumes with five people supposedly recently hired by a particular employer. When told that affirmative action had been used in recent hiring decisions, subjects paired the one African American in the employee group with the weakest resume at a rate significantly higher than did subjects in a second condition, where affirmative action was not mentioned.<sup>37</sup> In yet another study, subjects led to believe that affirmative action efforts had been involved in the influx of women into a particular occupation tended to assume that the women entering the field were unqualified. Researchers did not observe this denigrating effect in a control condition, with the same facts but no mention of affirmative action.<sup>38</sup> In all of these studies, the perception that employers used preferences in making

---

35. The terms "causal attribution," "attributional ambiguity," and "attribution bias" may require definition. Every day, people encounter events that require explanation. Cognitive psychologists refer to this explanatory process, which occurs without intention and with little awareness, as "causal attribution." See, e.g., SUSAN T. FISKE & SHELLEY E. TAYLOR, *SOCIAL COGNITION* 57-95 (1991). In many situations, events can be attributed to any of a number of different causes. This condition is termed "attributional ambiguity." Causal attribution is subject to a variety of systematic biases and distortions. These are referred to, collectively, as "attribution bias." *Id.* at 66-86.

36. See Luis T. Garcia et al., *The Effect of Affirmative Action on Attributions About Minority Group Members*, 49 J. OF PERSONALITY 427 (1981); Madeline E. Heilman et al., *Presumed Incompetent? Stigmatization and Affirmative Action Efforts*, 77 J. APPLIED PSYCHOL. 536 (1992); Rupert W. Nacoste, *Policy Schemas for Affirmative Action*, in *APPLICATIONS OF HEURISTICS AND BIASES TO SOCIAL ISSUES* 205 (Linda Heath et al. eds., 1994).

37. See Gregory B. Northcraft & Joanne Martin, *Double Jeopardy: Resistance to Affirmative Action from Potential Beneficiaries*, in *SEX ROLE STEREOTYPING AND AFFIRMATIVE ACTION POLICY* 81 (Barbara A. Gutek ed., 1982).

38. See Heilman & Herlihy, *supra* note 30.

selections negatively affected subjects' evaluations of beneficiaries' qualifications.

Another broad class of experiments investigated the effect of preferential selection on the evaluation of female co-workers. In one experiment, Jacobson and Koch paired male subjects with a female confederate, who was designated as the "leader" on a collaborative task.<sup>39</sup> Researchers told some subjects that they chose the female "leader" by chance. They told other subjects that she was selected because of her superior performance on a previously administered test of leadership ability. They told another group of subjects that the "leader" had been selected because of her sex. After subjects performed the task with the female "leader," researchers told them either that they had succeeded or that they had failed. In the "sex-based selection" condition, subjects blamed the female confederate for failure, but did not give her credit for their success, at a rate significantly higher than that observed in either the "chance" or "merit-based" selection conditions.<sup>40</sup>

Taken as a whole, this research suggests that in many contexts,<sup>41</sup> members of majority groups will assume that individual women and minorities selected in connection with preferential forms of affirmative action are less qualified and less capable than others. Some social scientists simply interpret this effect as a form of racism. Others explain it in somewhat less inflammatory and, I would suggest, ultimately more constructive terms, drawing on theories of causal attribution, stereotyping, and processes of stereotype change.

## 2. *Affirmative Action Preferences and the Perpetuation of Negative Stereotypes: The Role of Attributional Ambiguity*

As I described in earlier work, biases in the attribution of causation frequently result in discrimination against members of stereotyped outgroups.<sup>42</sup> Assume for example that two individuals, one a member of a stereotyped outgroup and the other a member of a dominant ingroup, experience the same negative outcome. If a member of the outgroup

39. See Marsha Jacobson & Walter Koch, *Women as Leaders: Performance Evaluation as a Function of Method of Leader Selection*, 20 ORGANIZATIONAL BEHAV. & HUM. PERF. 149 (1977).

40. See *id.* These findings were later replicated and extended by Madeline Heilman and her colleagues. See Madeline E. Heilman et al., *The Other Side of Affirmative Action: Reactions of Nonbeneficiaries to Sex-Based Preferential Selection*, 81 J. APPLIED PSYCHOL. 346 (1996).

41. The denigrating effect of perceived preferential selection, like its self-denigrating effects, appears to be context-sensitive, at least in some respects. So, for example, it has been shown to intensify where performance feedback is ambiguous, and to lessen where selection decisions are based both on group status and qualifications. See e.g., Madeline E. Heilman et al., *The Affirmative Action Stigma of Incompetence: Effects of Performance Information Ambiguity*, 40 ACAD. OF MGMT. J. 603 (1997).

42. See Linda Hamilton Krieger, *The Content of Our Categories: A Cognitive Bias Approach to Discrimination and Equal Employment Opportunity*, 47 STAN. L. REV. 1161 (1995).

experiences the negative outcome, and if it is consistent with a stereotype associated with that group, ingroup members will tend to attribute the bad outcome to stereotype-consistent traits, such as lack of ability. If, on the other hand, a member of a preferred ingroup experiences the negative outcome, it will more likely be attributed to transient external factors, such as bad luck or task difficulty.<sup>43</sup>

Favorable outcomes are attributed in a similar pattern. If a positive outcome contradicts a negative group stereotype, success by the stereotyped outgroup member is apt to be attributed to transient or external factors. "Expected success" by a dominant group member is more likely to be attributed to dispositional factors, like skill or hard work.<sup>44</sup>

Echoing the concerns of social commentators like Shelby Steele, social cognition theorist Brenda Major and her colleagues suggest that affirmative action preferences exacerbate attribution biases of this type.<sup>45</sup> Preferences create attributional ambiguity. For groups operating under a suspicion of inferiority, the presence of a preference provides a plausible reason for success in obtaining employment or admissions to prestigious educational programs. This plausible explanation, combined with preexisting negative expectancies relating to ability or achievement, lessens the likelihood that others will attribute beneficiaries' success to merit.

It is important to note that this negative effect depends not only on the presence of a preference, but also on the existence of negative stereotypes associated with the preferred group. This may explain the asymmetry between male and female subjects' reactions to receiving a sex-based preference.<sup>46</sup> As described earlier,<sup>47</sup> various researchers have found that women who believed that they had been selected because of their sex rated their competence relatively lower than did women who believed they had been selected based on merit. However, *male* subjects who were told that they had been selected because of their sex rated their own competence no differently than did male subjects who were

---

43. See, e.g., Galen V. Bodenhausen & Robert S. Wyer, Jr., *Effects of Stereotypes on Decision-making and Information-Processing Strategies*, 48 J. PERSONALITY & SOC. PSYCHOL. 267 (1985); Thomas F. Pettigrew, *The Ultimate Attribution Error: Extending Allport's Cognitive Analysis of Prejudice*, 5 PERSONALITY & SOC. PSYCHOL. BULL. 461 (1979).

44. See, e.g., Kay Deaux, *Sex: A Perspective on the Attribution Process*, in 1 NEW DIRECTIONS IN ATTRIBUTION RESEARCH 335 (John H. Harvey et al. eds., 1976); Madeline E. Heilman & Richard A. Guzzo, *The Perceived Cause of Work Success as a Mediator of Sex Discrimination in Organizations*, 21 ORGANIZATIONAL BEHAV. & HUM. PERF. 346 (1978).

45. See, e.g., Brenda Major et al., *Attributional Ambiguity of Affirmative Action*, 15 BASIC & APPLIED PSYCHOL. 113 (1994); Brenda Major & Jennifer Crocker, *Social Stigma: The Consequences of Attributional Ambiguity*, in AFFECT, COGNITION, AND STEREOTYPING 345 (Diane M. Mackie & David L. Hamilton eds., 1993).

46. See sources cited *supra* notes 21, 22.

47. See text accompanying notes 15-19.



told that their selection was merit based.<sup>48</sup> From these results, one can reasonably conclude that preferential selection *potentiates* pre-existing stereotypes associated with members of target groups, it does not create those stereotypes in the first instance.

Be that as it may, there are serious bases for concern that preferential forms of affirmative action might reinforce subtle negative expectancies relating to members of beneficiary groups. For example, using overt preferences in selection decision making can increase the salience of race or gender in those institutions. As many social cognition researchers have demonstrated, the more salient a particular characteristic in the mind of an observer, the more likely that observer will use it in making causal attributions.<sup>49</sup> Thus, we can expect that people will make race-or gender-relevant attributions more frequently in situations where race or gender is salient. Given that attributions are much more common following negative events than following positive events,<sup>50</sup> the increased salience of gender or racial characteristics occasioned by their overt use in decision making will increase the number of negative attributions to gender or race. This in turn will reinforce negative gender or racial stereotypes.<sup>51</sup>

Examination of this problem from a broader theoretical perspective only amplifies concern that preferential forms of affirmative action may reinforce negative gender and racial stereotypes, thus hampering the process of stereotype change. A fairly extensive body of theoretical and empirical work explores the mechanisms by which stereotypes persist in the face of disconfirming information. Three primary mechanisms have been identified: selective memory effects, expectancy-confirming attribution biases,<sup>52</sup> and a process known as "subtyping."<sup>53</sup>

---

48. See, Heilman et al., *supra* note 18; Heilman et al., *supra* note 15. Similar results have been observed with respect to the effect of gender-based selection on task preferences. When they believed they had been selected because of their gender, women, but not men, tended to avoid participation in more demanding tasks. See Heilman et al., *supra* note 17.

49. See generally Leslie Zebrowitz McArthur, *What Grabs You? The Role of Attention in Impression Formation and Causal Attribution*, in 1 *SOCIAL COGNITION* 201-06 (E. Tory Higgins et al. eds., 1981); Shelley E. Taylor & Susan T. Fiske, *Salience, Attention, and Attribution: Top of the Head Phenomena*, in 11 *ADVANCES IN EXPERIMENTAL SOCIAL PSYCHOLOGY* 249, 264-65 (Leonard Berkowitz ed., 1978).

50. See generally David L. Hamilton & Robert K. Gifford, *Illusory Correlation in Interpersonal Perception: A Cognitive Basis of Stereotypic Judgments*, 12 *J. EXPERIMENTAL SOC. PSYCHOL.* 392 (1976).

51. See Faye Crosby & Susan Clayton, *Affirmative Action and the Issue of Expectancies* 46 *I. SOC. ISSUES* 61, 68 (1990).

52. As described *supra* note 35, causal attribution is subject to a number of systematic cognitive biases. One of these, the expectancy confirmation effect, preserves stereotypes and other implicit theories from the corrosive effects of potentially disconfirming information. As described in more detail below, see *infra* Section II.A.2.a., by attributing stereotype relevant events to one as opposed to

With respect to selective memory effects, people readily recognize stereotype-consistent data as relevant. Once recognized, stereotype-consistent information tends to be encoded into and stored in memory in a form that makes it readily available for making stereotype-relevant social judgments.<sup>54</sup> This selective memory mechanism, explored later in this Article, provides a powerful critique of the "colorblindness" approach to nondiscrimination, but it is not particularly relevant to our present inquiry.

More relevant is a second phenomenon, known as the expectancy-confirming attribution bias.<sup>55</sup> During the early 1980s, Jennifer Crocker and her colleagues conducted a series of experiments attempting to explain an apparent contradiction between two robust research findings. One body of research had convincingly demonstrated that stereotypes and other forms of initial expectancies are highly resistant to subsequent expectancy-inconsistent information.<sup>56</sup> Another set of findings indicated that even though this incongruent information did not seem to alter impressions, it was particularly likely to be recalled.<sup>57</sup> Why, Crocker asked, would expectancy-inconsistent information be recalled at a relatively high rate and yet not alter the prior inconsistent expectancy?

Crocker and her collaborators explained this apparent contradiction by demonstrating that people tend to recall stereotype-

---

another cause, social perceivers may insulate their prior theories, including stereotypic expectancies, from the events' theory-disconfirming implications.

53. A category, including a social category such as "black," "woman," or "Hispanic," will become increasingly differentiated as a perceiver gains more familiarity with its constituents. Over time, categories develop a reverse tree-like structure, with the "superordinate" category (i.e. woman) at the top of the tree, and a proliferation of subordinate categories or "subtypes" branching out beneath it in a hierarchical organization. For a discussion of the nature and function of subtypes and their relationship to processes of categorization more generally, see, c.g., Patricia W. Linville et al., *Stereotyping and Perceived Distributions of Social Characteristics: An Application to Ingroup-Outgroup Perception*, in PREJUDICE, DISCRIMINATION, AND RACISM 165, 166 (John F. Dovidio & Samuel L. Gaertner eds., 1986) (defining "subtype"); see also Eleanor Rosch, *Principles of Categorization*, in COGNITION AND CATEGORIZATION 27, 30-35 (Eleanor Rosch & Barbara B. Lloyd eds., 1978) (describing the vertical organization of superordinate and subordinate categories).

54. See generally Naney Cantor & Walter Mischel, *Prototypicality and Personality: Effects on Free Recall and Personality Impressions*, 13 J. RES. IN PERSONALITY 187 (1979); Myron Rothbart, *Memory Processes and Social Beliefs*, in COGNITIVE PROCESSES IN STEREOTYPING AND INTERGROUP BEHAVIOR 145 (David L. Hamilton ed., 1978); Myron Rothbart et al., *Recall for Confirming Events: Memory Processes and the Maintenance of Social Stereotypes*, 15 J. EXPERIMENTAL SOC. PSYCHOL. 343 (1979).

55. See Jennifer Crocker et al., *Person Memory and Causal Attributions*, 44 J. PERSONALITY & SOC. PSYCHOL. 55 (1983).

56. See Crosby & Clayton, *supra* note 51.

57. See, e.g., Reid Hastie, *Memory for Behavioral Information That Confirms or Contradicts a Personality Impression*, in PERSON MEMORY 155, 174-75 (Reid Hastie et al. eds., 1980); Reid Hastie & Purohit Anand Kumar, *Person Memory: Personality Traits as Organizing Principles in Memory for Behaviors*, 37 J. PERSONALITY & SOC. PSYCHOL. 25, 26 (1979).

inconsistent information about a target individual at a higher than expected rate *only* when they have attributed that information to external, situational factors, rather than to internal dispositional factors relevant to the stereotype's accuracy. It is the increased attention and processing accorded inconsistent information during this causal attribution process that results in the observed high rate of recall. But the increased attention and processing also results in an attribution of the expectancy-inconsistent information that insulates the expectancy from the information's apparently disconfirming effect.<sup>58</sup> The information is, in a sense, "explained away," leaving the original expectancy undisturbed.

The implications of this model for the affirmative action debate are evident. Absent some explanatory theory, the presence of women or minorities in prestigious educational programs, nontraditional employment settings, or positions otherwise indicating high levels of ability and accomplishment would disconfirm a variety of inconsistent, negative racial or gender stereotypes. Over time, as such instances proliferated, they would be expected to erode the stereotypes they disconfirmed.<sup>59</sup> Preferential selection, however, provides a plausible situational attribution for the stereotype-inconsistent information. Once female or minority presence in "high places" is attributed to preferential selection rather than to merit-related factors, pre-existing negative stereotypes are insulated from the potentially disconfirming effect of the otherwise stereotype-disconfirming data.

Subtyping, a third mechanism involved in stereotype maintenance, presents one final reason for concern. The effect of stereotype-inconsistent evidence appears to depend on the frequency and distribution of its occurrence.<sup>60</sup> Specifically, if disconfirming information is concentrated in a single individual or in a small or otherwise discrete group, such individuals are "subtyped" and disassociated from the larger category with which they are otherwise associated. Severing the subtype thus preserves the schematic expectancies associated with the superordinate category as a whole.

The emergence of the stereotype "affirmative action black," frequently discussed in the aversive racism literature,<sup>61</sup> indicates that this

---

58. See Crocker et al., *supra* note 55, at 63-64. See also Reid Hastie, *Causes & Effects of Causal Attribution*, 46 J. PERSONALITY & SOC. PSYCHOL. 44, 53-55 (1984).

59. For a thorough elaboration of this model of stereotype change, see Marilyn B. Brewer & Roderick M. Kramer, *The Psychology of Intergroup Attitudes and Behavior*, 36 ANN. REV. PSYCHOL. 219, 221 (1985).

60. See *id.*

61. See, e.g., John F. Dovidio et al., *Resistance to Affirmative Action: The Implications of Aversive Racism*, in AFFIRMATIVE ACTION IN PERSPECTIVE 83 (Fletcher Blanchard & Faye Crosby eds., 1989).

subtyping process has occurred with respect to racial categorization and stereotyping. Once subtyped as "affirmative action beneficiaries," women or minorities in "unexpected" places lose their power to alter the negative schematic expectancies related to their social group.

In sum, attributional models of observers' and beneficiaries' responses to affirmative action raise concerns that preferential selection may indeed perpetuate negative stereotypes. However, these attributional models also suggest a methodological critique of the empirical evidence supporting this view.

As discussed earlier, numerous studies have shown that when members of stigmatized groups are selected to receive some valued outcome based on their group membership, they are more likely to respond negatively, devalue their own performance and the performance of others in their group, and devalue the domain itself, than are members of stigmatized groups selected on the basis of personal merit. But as Brenda Major and her colleagues have demonstrated, these effects weaken substantially when both group membership and personal qualifications play a role in selection decision making.<sup>62</sup>

One might reasonably argue that this defect in experimental design substantially circumscribes the applicability of empirical research on negative reactions to affirmative action. Affirmative action plans, be they private or government-instituted, voluntary or court-ordered, can never legally use group membership as the only factor in making selection decisions.<sup>63</sup> Group membership may be used only as one factor in choosing between otherwise qualified applicants. Given that the policies simulated in the affirmative action experiments would be unlawful in any context, one could argue that these studies have little utility in evaluating existing affirmative action policies.

On the other hand, one could argue that whether group status is used only in conjunction with other factors, what actually matters is the *perception* that selections are based on strong group preferences at the expense of merit-related factors. Thus, the discourse surrounding affirmative action will have more effect on public reactions to such policies than will the actual details of the programs themselves. To understand the precise effect that the discourse surrounding affirmative action might have on intergroup relations, one needs to know something about the psychology of procedural justice.

---

62. See Major et al., *supra* note 21, at 133.

63. See *Johnson v. Transportation Agency*, 480 U.S. 616, 622, 648-56 (1987) (Title VII); *United States v. Paradise*, 480 U.S. 149, 171 (1987); *Regents of the University of California v. Bakke*, 438 U.S. 265, 319-20 (1978) (Equal Protection Clause).

### C. *Preferences and Procedural Justice*

People's reactions to distributive choices are influenced not only by resulting patterns of distribution, but also by the perceived fairness of the procedures used in making them.<sup>64</sup> The fairer the policy, the more positive the evaluations.<sup>65</sup>

Drawing on these principles, various researchers have suggested that people react not only to the distribution of outcomes occasioned by affirmative action policies, but also to the procedures through which those policies are implemented. According to this perspective, using an affirmative action preference does not in and of itself stigmatize beneficiaries. Rather, selection using a procedure perceived as unfair imposes stigma and leads to other negative effects described in the relevant literature.<sup>66</sup>

Research conducted under this model explores the particular conditions under which people are more or less likely to respond either positively or negatively to affirmative action policies. Taken as a whole, this research suggests that people react more favorably to affirmative action distributions where both merit and group membership factor into distribution decision making;<sup>67</sup> where beneficiaries are described as qualified under applicable merit-based criteria;<sup>68</sup> where an historical pattern of discrimination against the preferred group is made salient in the description of the procedure;<sup>69</sup> where rigid numerical quotas are avoided;<sup>70</sup> and where preferences are applied in the distribution of training or educational opportunities rather than in post-training hiring

64. See generally E. ALLAN LIND & TOM R. TYLER, *THE SOCIAL PSYCHOLOGY OF PROCEDURAL JUSTICE* (1988); JOHN W. THIBAUT & LAURENS WALKER, *PROCEDURAL JUSTICE* (1975).

65. See generally Tom R. Tyler, *WHY PEOPLE OBEY THE LAW* (1990). See also E. Allan Lind et al., *Procedure and Outcome Effects on Reactions to Adjudicated Resolution of Conflicts of Interest*, 39 J. PERSONALITY & SOC. PSYCHOL. 643 (1980).

66. See generally Rupert Barnes Nacoste, *Sources of Stigma: Analyzing the Psychology of Affirmative Action*, 12 LAW & POL'Y 175 (1990); Rupert W. Nacoste & Darrin Lehman, *Procedural Stigma*, 17 REPRESENTATIVE RES. SOC. PSYCHOL. 25 (1987). See also TOM R. TYLER ET AL., *SOCIAL JUSTICE IN A DIVERSE SOCIETY* 53-54 (1997).

67. See Nacoste, *supra* note 26. The more weight given to group membership, the less fair the procedure is perceived to be. Decreasing assessments of fairness are, in turn, associated with increasingly negative reactions to the procedure and its beneficiaries. See Jerald Greenberg, *Reactions to Procedural Injustice in Payment Distributions: Do the Means Justify the Ends?*, 72 J. APPLIED PSYCHOL. 55 (1987); E. Allan Lind et al., *supra* note 65.

68. See Nacoste & Lehman, *supra* note 66; Rupert W. Nacoste, *But Do They Care About Fairness? The Dynamics of Preferential Treatment and Minority Interest*, 8 BASIC & APPLIED SOC. PSYCHOL. 177 (1987).

69. See Nacoste & Lehman, *supra* note 66. See also Susan D. Clayton & Sandra S. Tangri, *The Justice of Affirmative Action*, in *AFFIRMATIVE ACTION IN PERSPECTIVE* 177 (Fletcher A. Blanchard & Faye J. Crosby eds., 1989).

70. See James R. Kluegel & Eliot Smith, *Affirmative Action Attitudes: Effects of Self-Interest, Racial Affect and Stratification Beliefs on Whites' Views*, 61 SOCIAL FORCES 797 (1983).

or promotion.<sup>71</sup> In short, the manner in which an affirmative action program is framed and implemented influences subjective perceptions of procedural justice. Such perceptions, in large measure, determine attitudes towards the program and its beneficiaries.

But if these perceptions drive resistance to affirmative action, and if the various factors described above are indeed influencing such perceptions, how can one account for the rejection of affirmative action reflected in Proposition 209? To survive constitutional or statutory scrutiny, all affirmative action programs must possess the very characteristics identified in those studies as engendering support for affirmative action. Under applicable Supreme Court case law, group membership can be used only as one factor among many in selecting between otherwise qualified applicants.<sup>72</sup> Employers cannot consider group membership at all absent a showing of manifest imbalances in traditionally segregated job categories. State actors must further show that other methods of correcting those disparities have failed.<sup>73</sup> The use of preferences must be time limited, and must not "unnecessarily trammel the interests of non-preferred group members."<sup>74</sup> Strict quotas and consideration of group membership in making layoff decisions are explicitly prohibited.<sup>75</sup>

There are at least three possible explanations for the opposition to affirmative action expressed in the vote on Proposition 209. One plausible explanation draws on Meir Dan-Cohen's concept of "acoustic separation."<sup>76</sup> Dan-Cohen observes that a distinction can aptly be drawn between two types of legal rules. The first, decision rules, consist of those legal principles which guide judges in making legal decisions in specific cases. The second, which Dan-Cohen refers to as conduct rules, consists of corresponding normative principles which the public understands as binding rules of conduct.

Although these two categories of rules interrelate in a variety of significant ways, they are not coextensive. Decision rules are communicated to the general public, and are thereby absorbed into corresponding

---

71. See Stanley H. Jones & Stuart W. Cook, *The Influence of Attitude on Judgments of the Effectiveness of Alternative Social Policies*, 32 J. PERSONALITY & SOC. PSYCHOL. 767 (1975); Francine Tougas & France Veilleux, *Who Likes Affirmative Action: Attitudinal Processes Among Men and Women*, in AFFIRMATIVE ACTION IN PERSPECTIVE 111 (Fletcher A. Blanchard & Faye Crosby eds., 1989).

72. See authorities cited, *supra* note 63.

73. See, e.g., *United States v. Paradise*, 480 U.S. 149, 171 (1987).

74. *Id.* at 183.

75. See, e.g., *Bakke*, 438 U.S. at 305, 319-20 (discussing quotas); *Wygant v. Jackson Board of Education*, 476 U.S. 267, 283-84 (1986) (discussing layoff decisions).

76. See Meir Dan-Cohen, *Decision Rules and Conduct Rules: On Acoustic Separation in Criminal Law*, 97 HARV. L. REV. 625 (1984).

conduct rules, through a process of selective transmission.<sup>77</sup> As a result, the two sets of rules may differ. Thus in certain respects, conduct rules, what people *think* the law is, may differ from decision rules, the actual legal principles which judges apply in adjudicating cases.

As Dan-Cohen's model suggests, the voting public may well be unaware of the procedural prerequisites and restrictions imposed on affirmative action programs by constitutional and statutory decision rules. If true, the presence of restrictive provisions would simply not affect people's subjective assessments of the fairness of affirmative action.

It is also possible that the particular procedures associated with the use of group membership in the distribution of valued outcomes have less real world impact on attitudes towards affirmative action than theory and research on procedural justice predicts. Or, because the weighting given to group membership as compared to "merit-related" criteria is interpreted as either substantive or procedural in nature, procedural justice models of resistance to affirmative action may actually confound equity-based and procedure-based determinants of opposition or support.

Two more cynical explanations may lie at the root of this phenomenon. First, people may know that the law prohibits the "strong" use of group membership but believe that decision makers simply break the law and use preferences in legally impermissible ways.<sup>78</sup> Or, it may be that affirmative action opponents have so successfully dominated the discourse on the subject that people reflexively associate affirmative action with the use of strong group preferences in selection decision making.

Finally, as other theorists suggest, it may simply be that resistance to affirmative action programs is neither equity-based nor fairness-based, but is rather an expression of "aversive racism."<sup>79</sup> In any event, whether explained by reference to equity-oriented theories underlying attributional models, or by models centering on the psychology of

---

77. See *id.* at 635.

78. For a careful articulation of this view, supported by numerous specific examples, see STEPHAN THERNSTROM & ABIGAIL THERNSTROM, *AMERICA IN BLACK AND WHITE* 386-461 (1997).

79. See, e.g., Dovidio et al., *supra* note 61. The concept of "aversive racism," developed by Samuel Gaertner and John Dovidio, refers to a particular type of ambivalence stemming from a conflict between egalitarian values on the one hand and unacknowledged negative beliefs and feelings towards blacks on the other. "Aversive racists," according to Gaertner and Dovidio, are highly motivated to preserve a non-prejudiced self image. Thus, in interracial contexts they will avoid acting in ways that directly reflect racial animus. However, the negative component of their racial feelings and beliefs expresses in indirect and subtle ways, where a negative response towards blacks in general, or towards a particular black individual, can be rationalized on some basis other than race. See generally Samuel L. Gaertner & John F. Dovidio, *The Aversive Form of Racism*, in *PREJUDICE, DISCRIMINATION, AND RACISM* 61 (John F. Dovidio & Samuel L. Gaertner eds., 1986).

procedural justice, the results of numerous empirical studies present substantial ground for concern that preferential forms of affirmative action are associated with substantial, identifiable costs.

#### *D. Preferences and Social Identity*

Affirmative action opponents argue that one cannot use race, gender, or ethnicity as decision-relevant criteria in the allocation of scarce resources and at the same time build a society in which racial, gender, or ethnic categories become increasingly irrelevant in determining patterns of social perception, judgment, and behavior. As discussed above, theory and research in social cognition suggest that the use of group membership in selection decisions may indeed influence intergroup perception on an individual level. Research guided by social identity theory suggests that the use of group preferences may have effects on patterns of intergroup responding more broadly conceived.

A substantial body of work on intergroup behavior, representing over thirty years of empirical research, demonstrates that once divided into groups, even on a trivial or random basis, people exhibit powerful intergroup biases.<sup>80</sup> The mere introduction of "groupness" into a situation causes people to favor ingroup members in the allocation of rewards,<sup>81</sup> in the evaluation of performance,<sup>82</sup> in memory for positive versus negative behaviors,<sup>83</sup> and in the attribution of success or failure.<sup>84</sup> Increasing the salience of group distinctions exacerbates these effects, while decreasing group salience tends to erode them.<sup>85</sup>

The implications of this work for the debate over affirmative action are conflicting. On the one hand, both social identity theory and social identity research provide some evidence for the proposition that the preferential use of gender, race, or ethnicity in allocating competitive outcomes aggravates intergroup discrimination. For example, placing groups into competition has been shown to increase intergroup

---

80. For a lucid and thorough review of this literature, see Marilynn B. Brewer, *In-Group Bias in the Minimal Intergroup Situation: A Cognitive-Motivational Analysis*, 86 PSYCHOLOGICAL BULL. 307 (1989).

81. See Henri Tajfel et al., *Social Categorization and Intergroup Behavior*, 1 EUR. J. SOC. PSYCHOL. 149 (1971).

82. See C. K. Ferguson & H. H. Kelly, *Significant Factors in Over-evaluation of Own Group's Products*, 69 J. ABNORMAL & SOC. PSYCHOL. 223 (1964).

83. See John W. Howard & Myron Rothbart, *Social Categorization and Memory for In-Group and Out-Group Behavior*, 38 J. PERSONALITY & SOC. PSYCHOL. 301, 303-06 (1980).

84. See David A. Wilder, *Perceiving Persons as a Group: Categorization and Intergroup Relations*, in COGNITIVE PROCESSES IN STEREOTYPING AND INTERGROUP BEHAVIOR 213, 217 (David L. Hamilton ed., 1981).

85. See generally Samuel L. Gaertner et al., *Reducing Intergroup Bias: The Benefits of Recategorization*, 57 J. PERSONALITY & SOC. PSYCHOL. 239 (1989).



bias in the allocation of benefits<sup>86</sup> and in the evaluation of outgroup performance.<sup>87</sup> Compounding this cause for concern, other research indicates that an episode of personal failure, experienced under conditions in which group membership has been made salient, exacerbates negative reactions to members of the contextually relevant outgroup.<sup>88</sup> As long as categorical boundaries remain salient, providing people with individuating information about members of outgroups has little effect on the tendency to treat outgroup members as an undifferentiated mass.<sup>89</sup> Sharply defined category boundaries tend to suppress outgroup differentiation.<sup>90</sup>

Taken as a whole, research in social identity suggests that reducing the salience of intergroup boundaries can reduce intergroup bias, while increasing category salience can exacerbate it. It would be difficult to argue that the explicit use of race, gender, or ethnicity in allocating economic resources does not increase the salience of these categorical distinctions. Thus, if one accepts the claim supported by social identity research that increased salience worsens tendencies toward intergroup bias, it becomes difficult to discount the argument that affirmative action preferences may ultimately aggravate intergroup tensions.

On the other hand, social identity theory suggests that in certain other respects, affirmative action can produce conditions under which intergroup relations will improve. Factors other than competition increase the salience of racial, gender, or ethnic category boundaries. Particularly, categories will be more salient if they are characterized by convergent category boundaries and less salient if they are characterized by cross-cutting category boundaries.<sup>91</sup> Category boundaries can be said to be "cross-cutting" when alternative sources of status, identity, or other sources of interconnection cross-cut rather than correspond to category membership. For example, racial mixing in schools, universities, neighborhoods, or interdependent cooperative workgroups results in a greater likelihood that members of different racial groups will occupy ingroups constructed along social dimensions other than

---

86. See Marilynn B. Brewer & Roderick M. Kramer, *The Psychology of Intergroup Attitudes and Behavior*, 36 ANN. REV. PSYCHOL. 219, 227 (1985).

87. See Bernard M. Bass & George Duntzman, *Biases in the Evaluation of One's Own Group, Its Allies and Opponents*, 7 J. CONFLICT RESOLUTION 16 (1963); W. Doise et al., *An Experimental Investigation into the Formation of Intergroup Representations*, 2 EUR. J. SOC. PSYCHOL. 203 (1972).

88. See James R. Meindl & Melvin J. Lerner, *Exacerbation of Extreme Responses to an Outgroup*, 47 J. PERSONALITY & SOC. PSYCHOL. 71 (1984).

89. See Norman Miller & Marilynn B. Brewer, *Categorization Effects on Ingroup and Outgroup Perception*, in PREJUDICE, DISCRIMINATION, AND RACISM 209, 227-28 (John. F. Dovidio & Samuel L. Gaertner eds., 1986).

90. See *id.* at 228.

91. See, e.g., Marilynn B. Brewer & Norman Miller, *Beyond the Contact Hypothesis: Theoretical Perspectives on Desegregation*, in GROUPS IN CONTACT 281, 283 (Norman Miller and Marilynn B. Brewer eds., 1984).

race. These structures provide opportunities for recategorizing members of otherwise distinct social groups. This multiplication of potential categorical structures renders each less significant and thus less influential in intergroup perception, judgment, and behavior.<sup>92</sup>

Additionally, the presence of relatively equal numbers of members of different gender, racial, or ethnic categories within a social situation renders category distinctions less salient, whereas the presence of a clear minority makes category membership more salient.<sup>93</sup> Moreover, placing members of different social categories into situations involving cooperative interdependence and individuating social interactions also appears to reduce categorical responding.<sup>94</sup>

Preferential forms of affirmative action result in the relatively rapid integration of those institutions which use them. The value of this rapid integration, especially in educational or other settings where opportunities for individuating interaction between members of different groups are plentiful and a newly created sense of group identity is strong, can hardly be underestimated. Although placing members of different social groups into situations involving cooperative, individuating contact and the creation of cross-cutting sources of social identity may not always succeed in eliminating intergroup bias, it is unquestionably necessary to that end. In sum, the overt use of group membership in selection decision making may have certain costs, but the disadvantages associated with segregation are most certainly worse.

Taken as a whole, social cognition, procedural justice, and social identity theory can be used either to oppose or support preferential forms of affirmative action. But even if it were possible to determine whether the costs of preferences outweighed their benefits, such an exercise would hardly be worthwhile if it could be shown that, absent affirmative action, remaining policy tools were inadequate. To that issue our attention now turns.

## II

### AFTER AFFIRMATIVE ACTION:

#### RECOGNIZING DISCRIMINATION IN THE LAND OF THE COLORBLIND

What might we expect if every institution in the nation—every college and university, every corporation, every state and local public agency, and every arm and organ of the federal government—suddenly prohibited its employees from considering the race, sex, or national origin of applicants or employees in hiring, contracting, promotion, or admission to educational programs? What would happen if every

---

92. See *id.*; Gaertner et al., *supra* note 85, at 246.

93. See Brewer & Miller, *supra* note 91, at 290.

94. See *id.* at 290.

employment and admissions decision maker was told simply to "be colorblind," to base his or her decisions only on "considerations of merit"? Would they do it? Could they do it? Could we identify those who did not do it, whose decisions were tainted by intergroup bias?

The answers to these questions are, quite simply, "no," "no," and "no." Perhaps constitutions can be colorblind. Perhaps official government or corporate policies can be colorblind. But human beings living in a society in which history, ideology, law, and patterns of social, economic, and political distribution have made race, sex, and ethnicity salient, cannot be colorblind. The "colorblindness" approach to nondiscrimination will prove ineffective because it provides neither a framework for enabling people to recognize the effects of race, gender, or national origin on their perceptions and judgments, nor the tools required to help them counteract those effects. Indeed, a color blindness-centered interpretation of the nondiscrimination principle, coupled with well-meaning people's awareness that they *do* categorize along racial and ethnic lines, may exacerbate the very intergroup anxiety and ambivalence that lead to what social psychologists refer to as aversive racism.

Furthermore, decision makers cannot base selection decisions only on colorblind considerations of merit for the simple reason that merit has a color. Conceptions of merit are socially and politically constructed and are shaped by the same ingroup preferences that give rise to other subtle forms of intergroup bias. Affirmative action preferences have, in many ways, diverted our attention from the biases inherent in the construction of merit. But if preferences are eliminated, this problem and the inequities it generates will soon rise into sharp relief.

Finally, there is substantial reason to doubt that remaining law enforcement tools, particularly the adjudication of individual disparate treatment cases, will prove effective in identifying and remedying subtle but pervasive forms of intergroup bias. For a variety of reasons, reliance on individual disparate treatment adjudication can be expected to result in the serious underidentification of discrimination by judicial decision makers, victims, and private fact finders.

#### A. *The Inefficacy of Colorblindness as a Normative Construct*

As I have attempted to demonstrate in earlier work,<sup>95</sup> existing antidiscrimination law constructs intergroup bias as something that

---

95. In earlier work, I develop in extensive detail the claim that a substantial set of biased employment decisions derive not from discriminatory motive or intent, but from a variety of cognitive distortions that systematically bias intergroup perception and judgment even among the well-intentioned. See Krieger, *supra* note 42, at 1182-1211. Using more recent research, I intend to build upon rather than reiterate here the perspective taken in that earlier investigation.

occurs when a "discriminatory purpose" motivates a decision. In other words, in order for a decision to be considered "discriminatory" under a statute such as Title VII, the disparate treatment plaintiff must show that the employer chose to take the negative action against him because of his membership in a particular protected class. Thus, to say that discrimination is intentional means that the decision stands in a particular sort of close relation to the target person's group status. Specifically, in the decision maker's mental process, there must be some syllogistic connection between the two.

So, for example, existing antidiscrimination law understands cognitive stereotypes as causing discrimination through the operation of a conscious, syllogistic reasoning process, through which the decision maker uses a person's group status in the following sort of way:

Major Premise:	Women with young children are preoccupied with family responsibilities and do not put their jobs first;
Minor Premise:	This applicant is a woman with young children;
Conclusion:	This applicant cannot be expected to put the job first.

Current antidiscrimination law further conceives gender role expectations, or normative stereotypes, as causing discrimination through the operation of a similar sort of syllogistic reasoning:

Major Premise:	Women with young children <i>should be</i> preoccupied with family responsibilities and should not hold jobs that will compete with the responsibilities associated with raising children;
Minor Premise (1)	The rigors of this job can be expected to conflict with family responsibilities associated with raising young children;
Minor Premise (2)	This applicant is a mother with young children;
Conclusion:	This applicant should not hold this job.

According to the existing jurisprudential model of discrimination, personal animosity may also lead to discrimination through the operation of implicit syllogistic reasoning:

Major Premise:	Working with black coworkers makes me feel uncomfortable;
Minor Premise:	This applicant is black;
Conclusion:	Working with him would make me feel uncomfortable.

In each of these contexts, the decision maker's thinking moves directly through the target person's group status. So long as we understand discrimination as operating in this way, we can rely on a color-blindness model of nondiscrimination to function as an effective normative principle. A social decision maker can refrain from discriminating simply by refraining from any syllogistic use of the target person's group status, in other words, by being "colorblind."

But as I have also attempted to demonstrate,<sup>96</sup> not all discrimination is of this sort. Much discrimination has little connection with discriminatory motive or intent. This sort of discrimination occurs when an individual's group status subtly, even unconsciously, affects a decision makers' subjective perception of relevant traits, on which ostensibly non-discriminatory decision are subsequently based. This form of discrimination results from a variety of categorization-related cognitive biases, and can result in disparate treatment based on race, sex, national origin, or other factors, even among the well-intentioned.

I do not wish to rehash the evidence supporting this proposition described at length elsewhere. Rather, using related but more recent research, I wish to demonstrate that only "color-consciousness" can control these cognitive forms of intergroup bias.<sup>97</sup> This research strongly suggests that cognitive biases in social judgment operate automatically, without intention or awareness, and can be controlled only through subsequent, deliberate "mental correction" that takes group status squarely into account.<sup>98</sup>

### 1. *Automatic Processes in Intergroup Judgment*

In his early work on perceptual readiness,<sup>99</sup> Jerome Bruner observed that when a person receives information with the goal of forming an impression, his or her first cognitive task is to fit that information into some existing knowledge structure. As Bruner described, only when

---

96. See Krieger, *supra* note 42, at 1186-1201.

97. By "color consciousness," I mean the awareness that a person's race or other group status may be influencing one's perceptions or social judgments about that person.

98. By "mental correction," I mean a conscious, deliberate attempt to correct for the biasing effect of group status on social perception and judgment.

99. See generally Jerome S. Bruner, *On Perceptual Readiness*, 64 *PSYCHOLOGICAL REV.* 123 (1957).

behavioral information is encoded in this way does it becomes useful, or even meaningful.<sup>100</sup>

Of course, in many situations incoming information is ambiguous in that it is susceptible to varied interpretations. A student's volunteered but halting response to a question can be interpreted as reflecting dull-wittedness—or courageous engagement with a difficult subject. An employee's hesitancy in the face of an important decision may evince timidity—or prudence. As Bruner suggested, "All perception is generic,"<sup>101</sup> meaning that observed actions, like objects, take on meaning only when they are assigned to a particular trait construct—a pre-existing knowledge structure residing in the observer's mind.

In their attempts to understand this process more fully, cognitive psychologists originally assumed that assigning an action to a particular trait construct depended primarily on the extent of the "match" between the features of the action and those of the construct. Then in the late 1970s through the early 1980s, encouraged no doubt by Amos Tversky and Daniel Kahneman's seminal work on the availability heuristic,<sup>102</sup> various researchers began investigating the role of trait construct accessibility in social perception. Their work showed that the readiness with which a person will characterize a particular behavior in terms of any given trait construct is a function of that construct's availability in memory at the time the behavior is perceived.<sup>103</sup> Any activity, conscious or unconscious, that "primes" a particular trait construct will tend to increase its accessibility and the corresponding likelihood that ambiguous information will be assimilated or encoded in a manner consistent with that trait.<sup>104</sup>

---

100. See *id.* at 127.

101. *Id.* at 124.

102. See Amos Tversky & Daniel Kahneman, *Availability: A Heuristic for Judging Frequency and Probability*, 5 *COGNITIVE PSYCHOLOGY* 207 (1973).

103. See, e.g., E. Tory Higgins & Gillian King, *Accessibility of Social Constructs: Information-Processing Consequences of Individual and Contextual Variability*, in *PERSONALITY, COGNITION, AND SOCIAL INTERACTION* 69, 69-121 (Nancy Cantor & John F. Kihlstrom eds., 1981); E. Tory Higgins et al., *Category Accessibility and Impression Formation*, 13 *J. EXPERIMENTAL SOC. PSYCHOL.* 141 (1977); Thomas K. Srull & Robert S. Wyer, Jr., *Category Accessibility: Some Theoretical and Empirical Issues Concerning the Processing of Social Stimulus Information*, in 1 *SOCIAL COGNITION: THE ONTARIO SYMPOSIUM* 161-97 (E. Tory Higgins et al. eds., 1981); Robert S. Wyer, Jr. & Thomas K. Srull, *The Processing of Social Stimulus Information: A Conceptual Integration*, in *PERSON MEMORY* 227, 227-300 (Reid Hastie et al. eds., 1980).

104. See, e.g., John A. Bargh & Paula Pietromonaco, *Automatic Information Processing and Social Perception: The Influence of Trait Information Presented Outside of Conscious Awareness on Impression Formation*, 43 *J. PERSONALITY & SOC. PSYCHOL.* 437 (1982) (showing same effect following use of unconscious primes); John A. Bargh et al., *The Additive Nature of Chronic and Temporary Sources of Construct Accessibility*, 50 *J. PERSONALITY & SOC. PSYCHOL.* 869 (1986); Charles S. Carver et al., *Modeling: An Analysis in Terms of Category Accessibility*, 19 *J. EXPERIMENTAL SOC. PSYCHOL.* 403 (1983); Thomas K. Srull & Robert S. Wyer, Jr., *Category Accessibility and Social Perception: Some Implications for the Study of Person Memory and*

Social stereotypes bias perception in this general manner. As numerous researchers have demonstrated, one learns at an early age stereotypes of the major social groups in the United States.<sup>105</sup> These stereotypes have a long history of activation, and are likely to be highly accessible, regardless of whether they are believed.<sup>106</sup> They are invoked automatically when people encounter members of a stereotyped outgroup. Once activated, stereotypes serve to "prime" the trait constructs with which they are associated. Incoming behavioral information, especially if capable of various interpretations, is accordingly assimilated into those traits associated with the stereotype.<sup>107</sup>

This tendency to assimilate ambiguous information into stereotypic trait constructs might not be so serious if people were aware that they were doing it. To understand the significance of this process and its implications for the debate over affirmative action, it is useful to understand a phenomenon which attribution theorists refer to as spontaneous trait inference.

People are highly concerned with understanding why things happen in their social environments. Rightly or wrongly, we assume that understanding why something has happened will improve our power to predict or even control what will happen in the future.<sup>108</sup> To the extent that personality traits play an important role in understanding people's actions, one might expect the process of translating observed behaviors into trait-related meanings to occur with great frequency. Given that

---

*Interpersonal Judgments*, 38 J. PERSONALITY & SOC. PSYCHOL. 841 (1980) (demonstrating influence on interpretation of ambiguous behavior using conscious primes); Thomas K. Srull and Robert S. Wyer, Jr., *The Role of Category Accessibility in the Interpretation of Information About Persons: Some Determinants and Implications*, 37 J. PERSONALITY & SOC. PSYCHOL. 1660 (1979).

105. See generally Patricia G. Devine, *Automatic and Controlled Processes in Prejudice: The Role of Stereotypes and Personal Beliefs*, in ATTITUDE STRUCTURE & FUNCTION 181, 182-84 (Anthony R. Pratkanis et al. eds., 1989). See also Mary Ellen Goodman, RACE AWARENESS IN YOUNG CHILDREN (rev. ed. 1964); Phyllis A. Katz, *The Acquisition of Racial Attitudes in Children*, in TOWARDS THE ELIMINATION OF RACISM 125 (Phyllis A. Katz ed., 1976). In discussing her finding that, by the age of four, children have already begun to internalize the stereotypes of major social groups, Dr. Katz tells a chilling story. While at a grocery store in one town where she was conducting field research, Dr. Katz observed a three year old white child who, upon seeing an African-American infant in a stroller remarked, "Look mom, a baby maid!" See *id.* at 147.

106. See Patricia G. Devine, *Stereotypes and Prejudice: Their Automatic and Controlled Components*, 56 J. PERSONALITY & SOC. PSYCHOL. 5, 6-7 (1989); John F. Dovidio et al., *Racial Stereotypes: The Contents of Their Cognitive Representations*, 22 J. EXPERIMENTAL SOC. PSYCHOL. 22 (1986); Samuel L. Gaertner & John P. McLaughlin, *Racial Stereotypes: Associations and Ascriptions of Positive and Negative Characteristics*, 46 SOC. PSYCHOL. Q. 23 (1983). This tendency to interpret ambiguous behaviors in a manner consistent with ethnic stereotypes appears to develop quite early. See Krieger, *supra* note 42, at 1202-03.

107. See Higgins & King, *supra* note 103, at 72-75.

108. The earliest attribution theorists recognized this impulse. See, e.g., Fritz Heider, *THE PSYCHOLOGY OF INTERPERSONAL RELATIONS* (1958); Harold H. Kelly, *Attribution in Social Interaction*, in ATTRIBUTION: PERCEIVING THE CAUSES OF BEHAVIOR 1, 22 (Edward E. Jones et al. eds., 1972).

social perception and judgment processes become increasingly efficient with repeated execution, one might further hypothesize that trait inference processes could become so overlearned as to operate without intention or awareness, much like the processes involved in recognizing a word or a face.

This hypothesis appears to be correct. In a number of studies replicated in a variety of contexts by other researchers, New York University psychologist James S. Uleman and his colleagues demonstrated that European-American<sup>109</sup> subjects spontaneously encode behaviors into stable trait constructs without intention or awareness.<sup>110</sup> Thus, there is a strong tendency, at least among European Americans, to attribute stable dispositional qualities spontaneously, as part of the process of perceiving and encoding information about another person's behavior. To say that trait inference is "spontaneous" however is not quite the same as saying it is "automatic,"<sup>111</sup> and the difference is critical to equal opportunity policy.

Through the early-1980s, it was generally believed that a particular mental process was either entirely automatic or entirely deliberate.<sup>112</sup> Over time, however, cognitive processes came to be understood as falling along a continuum. On one side of that continuum lie "controlled processes," which require substantial processing capacity and occur with greater levels of focus and awareness. On the other side lie fully automatic processes, which occur without intention<sup>113</sup> or awareness, are difficult if not impossible to control once triggered, and

---

109. The tendency spontaneously to infer stable dispositional traits from observed behaviors is, at least in part, culturally determined. Cross-cultural studies have demonstrated that non-European-American subjects use fewer dispositional qualities to explain other's behaviors, and are more likely to locate the cause of behaviors in the situation confronting the actor. For more on these findings, see generally Joan G. Miller, *Culture and the Development of Everyday Social Explanation*, 46 J. PERSONALITY & SOC. PSYCHOL. 961 (1984); Leonard S. Newman, *Why Are Traits Inferred Spontaneously? A Developmental Approach*, 9 SOC. COGNITION 221, 247 (1991); E. Rhee et al., *Spontaneous Self-Descriptions and Ethnic Identities in Individualistic and Collectivistic Cultures*, 69 J. PERSONALITY & SOC. PSYCHOL. 142, 142-52 (1995). For a comprehensive overview of research on spontaneous trait inference, see James S. Uleman et al., *People as Flexible Interpreters: Evidence and Issues From Spontaneous Trait Inference*, 28 ADVANCES IN EXPERIMENTAL SOC. PSYCHOL. 211 (1996).

110. See generally Uleman et al., *supra* note 109 (reviewing the experimental evidence for and describing spontaneous trait inference).

111. Truly "automatic" mental processes possess four qualities: they are unintentional, they occur outside of awareness, they are efficient in their use of attentional resources, and, most importantly, they are uncontrollable. See generally John A. Bargh, *The Four Horsemen of Automaticity: Awareness, Intention, Efficiency and Control in Social Cognition*, in 1 HANDBOOK OF SOCIAL COGNITION 1, 1-2 (Robert S. Wyer, Jr. & Thomas K. Srull eds., 2d ed. 1994).

112. See *id.* at 2.

113. In the relevant literature, "intention" is defined as whether one is in control of the instigation of a particular mental process. So for example, the automatic process by which a wooden structure with four legs, a vertical back, a platform and horizontal side supports is categorized as a chair can be understood as occurring without intention.



interfere little with other ongoing mental activity.<sup>114</sup> Eventually, it became apparent that complex mental processes such as causal attribution and other forms of social inference were neither exclusively automatic nor exclusively controlled, but rather combined aspects of both.<sup>115</sup> Specifically, social inference came to be understood as comprising a chain of three sequential subprocesses: "categorization," in which the person perceived is identified and placed within an existing categorical structure; "characterization," in which spontaneous dispositional inferences are drawn from the observed behavior; and "correction," in which those dispositional inferences are adjusted to account for situational factors. While categorization and characterization are automatic, correction is controlled. It requires deliberate, effortful mental processing and will compete for cognitive resources with other information processing demands.<sup>116</sup>

As I have indicated, the recognition that social inference comprises both automatic and controlled processes bears significant implications for the colorblindness approach to non-discrimination. When we encounter a person in our social environment, we *automatically* place that person into pre-existing categorical structures.<sup>117</sup> It is virtually impossible to escape this proclivity. As Eleanor Rosch observed, the particular categories into which people will divide their environment are determined by how well those divisions help them understand that environment and predict future events.<sup>118</sup> Admonitions to refrain from categorizing the social environment in a particular way will prove ineffective if those categories reflect patterns made salient by history, culture, or observable patterns of economic, demographic, or political distribution. Whether we like it or not, it is highly implausible to assume that an American of one race encountering an American of another race would *not* notice racial attributes or use those attributes in initially categorizing the person perceived. It is one thing to maintain as a

---

114. See Laraine Winter et al., *How Automatic Are Social Judgments?* 49 J. PERSONALITY & Soc. PSYCHOL. 904, 905 (1985).

115. For an early challenge to the original understanding of automatic and controlled processing, see George A. Quattrone, *Overattribution and Unit Formation: When Behavior Engulfs the Person*, 42 J. PERSONALITY & Soc. PSYCHOL. 593 (1982). For a comprehensive discussion of how trait inference eventually came to be viewed as possessing some but not all characteristics of automatic processing, see, e.g., Bargh, *supra* note 111, at 2-3.

116. See Bargh, *supra* note 111, at 28-30; Daniel T. Gilbert, *Thinking Lightly About Others: Automatic Components of the Social Inference Process*, in UNINTENDED THOUGHT 189, 194 (James L. Uleman & John A. Bargh eds., 1989).

117. See Bargh, *supra* note 111, at 13; Marilyn B. Brewer, *A Dual Process Model of Impression Formation*, in 1 ADVANCES IN SOCIAL COGNITION 1, 3-9 (Thomas K. Srull & Robert S. Wyer, Jr. eds., 1988).

118. See Eleanor Rosch, *Human Categorization*, in 1 STUDIES IN CROSS-CULTURAL PSYCHOL. 1, 40 (Neil Warren ed., 1977).

normative matter that people should not categorize by race. But it is counterfactual to maintain that they will not do so.<sup>119</sup>

There is substantial theoretical and empirical support for the view that the presence of members or symbolic representations of a stereotyped group automatically activates stereotypes associated with that group.<sup>120</sup> As an automatic process, stereotype activation is unintentional—a person has no control over its initiation.<sup>121</sup> Once activated, the stereotype functions as a prime, thereby pulling spontaneous trait inference in a stereotype-consistent direction. As Patricia Devine demonstrated, this automatic process of stereotype activation will occur in both high and low prejudice individuals.<sup>122</sup>

Stereotypes are not the only or even the most insidious source of inference-biasing primes. In a fairly chilling series of studies, social cognition researchers have shown that merely thinking about one's self or one's ingroups, or being subliminally exposed to such ingroup primes such as "us" or "we," results in the automatic activation of generally more positive trait concepts.<sup>123</sup> Thus, it is reasonable to believe that people will interpret ambiguous behaviors of persons with whom they identify more positively than they will interpret similar behaviors of persons from whom they feel socially distant.

The significance of these processes and their implications for the colorblindness approach to nondiscrimination can hardly be overemphasized. Very little information shapes a social perceiver's impression of a target person. Rather, it is the perceiver's interpretation of the raw information that influences social judgment. If the target's social group membership influences these interpretations, and if one is unaware of the effect of such status on those interpretations, how can we expect a colorblindness approach to nondiscrimination to function successfully as a normative principle? Given the realities of social perception, we can anticipate that similarly situated people will be treated differently based on their group membership, because decision makers, influenced by these subtle forms of intergroup bias, will not perceive them as

---

119. See, e.g., Shelley E. Taylor et al., *Categorical and Contextual Bases of Person Memory and Stereotyping*, 36 J. PERSONALITY & SOC. PSYCHOL. 778 (1978) (describing how race is used as an encoding strategy irrespective of perceiver "intention").

120. See Brewer, *supra* note 111, at 3-9; Devine, *supra* note 105; Charles W. Perdue & Michael B. Gurtman, *Evidence for the Automaticity of Ageism*, 26 J. EXPERIMENTAL SOC. PSYCHOL. 199 (1990); Felicia Pratto & John A. Bargh, *Stereotyping Based on Apparently Individuating Information: Trait and Global Components of Sex Stereotypes Under Attention Overload*, 27 J. EXPERIMENTAL SOC. PSYCHOL. 27 (1991).

121. See Brewer, *supra* note 111, at 6; Devine, *supra* note 106, at 6, 89-90.

122. See Devine, *supra* note 106, at 12, 15.

123. See Charles W. Perdue et al., *Us and Them: Social Categorization and the Process of Intergroup Bias*, 59 J. PERSONALITY & SOC. PSYCHOL. 475, 482-83 (1990). See also Jennifer Crocker & Ian Schwartz, *Prejudice and Ingroup Favoritism in a Minimal Intergroup Situation: Effects of Self-Esteem*, 11 PERSONALITY & SOC. PSYCHOL. BULL. 379 (1985).

similarly situated at all. Nothing in the colorblindness approach to non-discrimination provides social decision makers with the tools required to recognize or to correct for biases of this sort.

According to spontaneous trait inference theory, only the application of deliberate, controlled, corrective processes can prevent stereotypes and subtle ingroup priming valences from biasing interpersonal judgment. As social cognition researchers Patricia Devine and Susan Fiske observe, it is neither that nonprejudiced individuals "do not notice" such traits as gender or ethnicity nor that the presence of a member of another group does not "prime" the stereotypes associated with those groups. Rather, insofar as cognitive sources of bias are concerned, the difference between people who discriminate and those who do not is that members of the latter group *notice* the influences of stereotypes on their thinking and counteract those influences by consciously adjusting responses in a nonprejudiced direction. This process, however, is effortful: it requires both strong motivation and a great deal of capacity, attention, and practice.<sup>124</sup> In short, controlling the biases stemming from such processes as spontaneous trait inference is substantially more complicated than it might at first seem.

## 2. *Controlled Processes and Nondiscrimination: Taming the Beast of Automaticity*

The major sources of error in human judgment divide into two broad types. Errors of the first type stem from a failure to know or apply normative rules of inference. Errors of the second type result from a phenomenon which Timothy Wilson and Nancy Brekke refer to as "mental contamination."<sup>125</sup> Mental contamination occurs when a person's judgment or behavior is corrupted by unconscious or uncontrollable mental processes which she would rather *not* have influence her behavior or decisions.<sup>126</sup> Judgment errors deriving from rule ignorance or incorrect rule application are easier to remedy than those resulting from mental contamination. Normative rules of inference, like the rule of regression to the mean,<sup>127</sup> supply specific procedures for solving the

---

124. See Devine, *supra* note 105, at 188.

125. See Timothy D. Wilson & Nancy Brekke, *Mental Contamination and Mental Correction: Unwanted Influences on Judgments and Evaluations*, 116 PSYCHOLOGICAL BULL. 117, 117 (1994).

126. See *id.*

127. Regression to the mean is a statistical phenomenon in which unusually high or low values in a distribution tend to be followed by more average values. So, for example, parents who are particularly tall tend to have children who are closer to average height; extremely high test scores tend to be followed by scores closer to the mean; and extremely high jury verdicts tend to be followed by lower ones that more closely approximate the average award for cases of that type. The tendency to overlook regression to the mean, often referred to as nonregressive prediction, leads to systematic

problems to which they pertain and can be consciously learned and deliberately applied.<sup>128</sup>

Correcting judgmental errors resulting from mental contamination is more difficult, in large part because simply teaching people a particular decision rule is unlikely to control biases of which they are unaware. This will be particularly true if the biases in question are difficult to recognize, or easily mistaken for valid, decision-relevant considerations. A supervisor evaluating employees for promotion, or a professor considering which student to hire as a research assistant, may know the rule, "don't discriminate against the black guy." But that kind of rule, in this case the colorblindness rule, cannot be applied in the same way as cost-benefit analysis or the rule of regression to the mean. It can be applied to eliminate facial discrimination—that is, a conscious, explicit policy of excluding a certain group of persons from consideration. It might even be applied to eliminate the *conscious, deliberate* use of group status as a proxy for decision-relevant traits like initiative or writing ability. But it cannot be applied to prevent or correct biases caused by emotional discomfort, the subconscious effects of stereotypes, causal attribution, or spontaneous trait inference, because it fails to provide a specific set of procedures or techniques which can be applied to the evaluation or decision task at hand. A normative decision rule, such as one prohibiting discrimination on the basis of race, cannot be applied to eliminate a source of bias if the decision maker is unaware that her judgment might be biased or is unable to control the effects of such bias for lack of applicable remedial tools.

As Wilson and Brekke explain, four discrete conditions must be satisfied if people are to control the effects of nonconscious biases. First, one must become aware of the nature of the particular mental process which threatens to bias one's judgment. Second, one must be motivated to correct its unwanted influence once it has been recognized. Third, one must be able to discern the direction and magnitude of the bias, lest it be "overcorrected" and judgment skewed in the opposite direction. And finally, one must have sufficient control over his or her mental processes to correct the effect of the unwanted influences.<sup>129</sup>

---

errors in judgment. See SCOTT PLOUS, *THE PSYCHOLOGY OF JUDGMENT AND DECISION-MAKING* 116-118 (1993).

128. It is well-established that people are able to learn these rules, apply them in appropriate contexts, and thereby reduce the incidence of judgment error in those situations to which they pertain. See, e.g., Geoffrey T. Fong et al., *The Effects of Statistical Training on Thinking About Everyday Problems*, 18 *COGNITIVE PSYCHOL.* 253 (1986); Geoffrey T. Fong & Richard E. Nisbett, *Immediate and Delayed Transfer of Training Effects in Statistical Reasoning*, 120 *J. EXPERIMENTAL PSYCHOL.: GENERAL* 34 (1991); Richard P. Larrick et al., *Teaching the Use of Cost-Benefit Reasoning in Everyday Life*, 1 *PSYCHOLOGICAL SCIENCE* 362 (1990); R. E. Nisbett et al., *Teaching Reasoning*, 238 *SCIENCE* 625 (1987).

129. See Wilson & Brekke, *supra* note 125, at 119-20.

While it is beyond the scope of this Article to review all of the problems encountered at each of these four stages, those with the most serious implications for the colorblindness model of nondiscrimination warrant attention here.

*a. Unawareness of Mental Process*

During the fall of 1996, while the campaign on Proposition 209 swirled around the University, I was teaching a class on employment discrimination law. Every day, I called on a different student to respond to questions about the cases prepared for that day's class. Most of my students were Caucasian; only three of sixty-five were African American. Let us assume for purposes of illustration, that one day I had called on one of these three African-American students, and that he experienced a certain amount of difficulty answering various questions. He was not unprepared, but many of his answers were halting and somewhat confused, leaving me with the initial impression that he was not particularly capable.

I flatly reject the belief that African-American law students are less intelligent than others. If asked at the beginning of the semester to predict how any one of my three African-American students would perform relative to their classmates, I would have vigorously objected to making a prediction in the absence of individuating information. Given that I am familiar with certain normative rules of inference such as the principle of regression to the mean, I would probably, if pushed, have predicted that his or her performance would be about average.

But it is also true that, although I reject them as untrue, I am *aware* of the stereotypes associated with intelligence, academic achievement, and African-American males. I too was exposed to those stereotypes at a very early age, before I developed my own powers of critical and moral intelligence and made a conscious decision to reject these stereotypes as inaccurate and unfair. But my nonprejudiced beliefs did not *displace* the stereotypes, which exist alongside and function independently of these beliefs. The stereotypes are triggered whether I believe in them or not.<sup>130</sup> And, once triggered, those stereotypes prime the trait constructs associated with them, constructs like "not too bright" or "underachieving," rather than "grappling courageously with a difficult subject."

So, what should one conclude if I had taken from this hypothetical encounter the initial impression that the student in question was "not particularly capable?" His performance had not been very good. Why

---

130. For further elaboration of the nature and empirical evidence supporting this "dual process model," see, e.g., Devine, *supra* note 105, at 186-88, and sources cited therein.

then, should I even question whether the student's race had anything to do with my judgment?

If such a situation were to arise, I would likely question my initial impression only if: (1) I was aware of the possibility that negative stereotypes of African-American males, or some other aspects of the situation had subtly influenced my judgment; and (2) I was motivated to do something about it.

Of course, other subtle cognitive sources of bias besides racial stereotypes could have "contaminated" my judgment as well. If I approached the issue mindfully, I would probably have noticed that over the course of the semester, a number of students I had called on had performed relatively poorly, some just as poorly as the student in question. But try as I might, if I had not made contemporaneous notes, I probably could not now remember just who those students were. I would remember the African-American student, but I would most likely have forgotten the others.

This scenario illustrates a common unconscious source of bias—the polarized evaluation of distinctive members of an otherwise largely homogeneous group.<sup>131</sup> It is well-established that people pay particularly close attention to distinctive stimulus objects, such as a "token" woman or minority group member. And the more attention we pay to something, the more about it we perceive, encode, and store in memory. Indeed, under conditions of high attention, we are more likely to encode an event visually, which makes it more readily available in memory and more influential in the formation of subsequent judgments. Accordingly, the poor performance of a distinctive minority student is more likely to be remembered, and will tend to be charged with a more powerful negative valence, than the poor performance of a majority white student.

If I were unschooled in these sorts of salience or expectancy-related biases, I would likely remain unaware that the student's race had played any role in the formation of my initial impression that he was "not particularly capable." But could I fairly deny that he had been negatively judged, at least in part, because of his race?

The colorblindness approach to nondiscrimination is dangerous because it leads a decision maker to believe that, so long as she is not consciously thinking about race, she is not discriminating. But social cognition theory teaches that, in a culture pervaded by racial stereotypes, or where persons of one race constitute a small minority in an otherwise homogeneous group, one *must* think about race in order *not* to discriminate. In short, the colorblindness principle discourages the first step prerequisite to controlling cognitive sources of intergroup bias.

---

131. For a more thorough discussion of this phenomenon, see Krieger, *supra* note 42, at 1193-95.

*b. The Role of Motivation*

All "dual process"<sup>132</sup> models of social inference posit that in order to correct errors caused by an automatic mental process, people must not only be aware of the process, but must also be motivated to control its biasing effects. Of course, the development of awareness itself requires motivation. Despite early controversy on particular issues,<sup>133</sup> it is now relatively well-accepted that people lack awareness of a large proportion of mental processing,<sup>134</sup> including the processes comprising impression formation.<sup>135</sup>

Developing self-awareness of such processes and correcting for the various biases inherent in them is in any context objectively difficult. Even after they are activated, the controlled processes required for mental correction require a great deal of capacity and attention and will compete for cognitive resources with other mental demands. But in the context of intergroup discrimination, the order is taller still. Increasing awareness of and sustained attention to the biasing effects of racial, ethnic, or gender stereotypes on one's social judgments is apt to engender fear of moral opprobrium and substantial psychological discomfort. Thus, especially in the context of reducing intergroup bias, there is little reason to assume that people will expend the effort or bear the psychological discomfort associated with mental correction unless they have strong motivations for so doing. Thus, the colorblindness approach not only fails to provide incentives for developing an awareness of mental contamination, but the model itself and the rhetoric that often accompanies it actually establish *disincentives* for so doing.

Affirmative action opponents, including racial minorities who have benefited from those preferences, frequently argue that it is time for minorities to stop blaming their problems on discrimination.<sup>136</sup> There are undoubtedly disadvantages associated with "victim mentalities," and there is ample reason to air those disadvantages in a broad-ranging

---

132. I borrow this term from Marilyn B. Brewer, *see* Brewer, *supra* note 118, but use it here to refer as well to James Uleman's model of spontaneous trait inference, Patricia Devine's disassociation model of stereotyping and prejudice, and Wilson & Brekke's model of mental contamination and correction. *See* Devine, *supra* note 106; Uleman, *supra* note 109; Wilson & Brekke, *supra* note 125.

133. For a discussion of these controversies, *see* Timothy D. Wilson & Julie I. Stone, *Limitations on Self-Knowledge: More on Telling More Than We Can Know*, in *SELF, SITUATIONS, & SOCIAL BEHAVIOR* 167 (Philip Shaver ed., 1985).

134. *See generally* J.F. Kihlstrom, *The Cognitive Unconscious*, 237 *SCIENCE* 1445 (1987); PAWEŁ LEWICKI, *NONCONSCIOUS SOCIAL INFORMATION PROCESSING* (1986); JAMES S. ULEMAN & JOHN A. BARGH, *UNINTENDED THOUGHT* (1989).

135. *See* Gilbert, *supra* note 116, at 190. *See generally* Ellen J. Langer, *Rethinking the Role of Thought in Social Interaction*, in 2 *NEW DIRECTIONS IN ATTRIBUTION RESEARCH* 35 (John H. Harvey et al. eds., 1978); LEWICKI, *supra* note 134.

136. *See, e.g.,* STEELE, *supra* note 13, at 118-19; Thomas Sowell, *Black Progress Can't Be Legislated*, *WASH. POST*, Aug. 12, 1984, at B1.

policy debate. But one must remember that the statement that "it is time for minorities to stop blaming their problems on discrimination" is heard not only by members of minority groups, but by members of the majority as well. To them, it sends the message, "You can stop worrying about discrimination now."

In such a rhetorical environment, people, most of whom are predisposed to overconfidence in their capacity to resist cognitive biases, are unlikely to expend the effort and endure the discomfort required to cultivate awareness of the subtle effects of race on their perception and judgment. If people equate "not thinking about race" with nondiscrimination, neither their own desire to live true to some inner set of egalitarian values, nor their desire to conform to antidiscrimination norms will help them recognize or control cognitive forms of intergroup bias. If they do seek to engage in mental correction, adjusting for the effects of cognitive biases will feel like racial preferencing. In other words, the very things required to eliminate cognitive forms of bias will be experienced as a violation of the colorblindness norm.

*c. The Limits of Mental Correction*

As Timothy Wilson and Nancy Brekke observe, even if a person becomes aware that some unwanted mental process has tainted her judgment, she may not be able to determine the magnitude of the resulting bias.<sup>137</sup> For example, in my hypothetical interaction with my African-American student, I might have become aware that stereotype or salience-related biases had influenced my assessment of his performance. But assuming that I were eventually required to formulate an evaluation, how far, if at all, should I adjust it? There is really no way for me to assess how much of my impression is fairly attributable to bias and how much to the student's flawed performance.

Even more troubling is the question whether "correction" is feasible at all. Once the initial impression that the student was "not particularly capable" had been formed, would I be able to erase it from my mind, or prevent it from influencing my impressions of him in connection with future interactions?

There is ample reason to fear that I would not. In a series of now classic studies, Stanford psychologist Lee Ross and his colleagues demonstrated that, even after a belief is discredited, the causal explanations generated to support it persist, giving the discredited belief a kind of cognitive life after death.<sup>138</sup> More recently, University of Texas

---

137. See Wilson & Brekke, *supra* note 125, at 120.

138. See Craig A. Anderson et al., *Perseverance of Social Theories: The Role of Explanation in the Persistence of Discredited Information*, 39 J. PERSONALITY & SOC. PSYCHOL. 1037 (1980);



psychologists Daniel Gilbert and Randall Osborne extended these observations to the process of spontaneous trait inference.<sup>139</sup> Their work demonstrates that once a trait inference is made, subsequent efforts to adjust it may prove ineffective. As they observed, misperceptions are "metastatic."<sup>140</sup> Controlled processing may correct the original misperception, but it often fails to eliminate subsidiary changes that the original misperception engendered. These endure and influence subsequent judgments of the person perceived. It is easier to forbear from action based on a biased impression than to eliminate the impression itself.<sup>141</sup>

Assuming I realized that my impression formation process was potentially biased, I might have decided to reject the view that the student was not particularly capable. I might even have decided not to take action based on my initial impression of his in-class performance, for example, deciding not to use it in calculating his grade. But I would probably not be able to erase the impression from my mind. Given its enduring presence, I would likely experience any adjustments in my subsequent behavior toward or expressed beliefs about the student as a form of racial preferencing. "After all," I might tell myself, "if the student weren't African-American, I wouldn't be bending over backwards like this." What the colorblindness perspective would allow me to forget, or never teach me in the first place, is that if the student were not African-American, I probably would not have remembered his performance at all.

### B. Merit

In the debate over affirmative action, preferential selection is frequently contrasted with selection based on merit. Without question, the problem of merit is central to the debate. Although affirmative action supporters have attempted to sidestep the issue by comparing affirmative action to veterans' preferences or legacy admissions, it would be unwise to ignore the challenge presented by the normative principle of merit-based decision making.

#### 1. The Merit Principle and Perceptions of Procedural Justice

At least in the United States, virtually every psychological model advanced to explain people's responses to distributive allocations reflects the centrality of merit in people's perceptions of fairness.

---

Lee Ross et al., *Perseverance in Self-Perception and Social Perception: Biased Attributional Processes in the Debriefing Paradigm*, 32 J. PERSONALITY & SOC. PSYCHOL. 880 (1975).

139. See Daniel T. Gilbert & Randall E. Osborne, *Thinking Backward: Some Curable and Incurable Consequences of Cognitive Busyness*, 57 J. PERSONALITY & SOC. PSYCHOL. 940 (1989).

140. *Id.* at 946.

141. See *id.*

Equity theory, for example, observes that distributive outcomes are perceived as just to the extent that contributions correspond with rewards.<sup>142</sup> When individual "inputs" and "outputs" fail to correspond, the over- and under-rewarded, the directly affected, and the third-party observer respond negatively.<sup>143</sup> And while the equity principle is not the only justice rule influencing subjective responses to distributive allocations,<sup>144</sup> it appears to be the one most commonly applied in the workplace.<sup>145</sup>

Merit-based decision-making principles also figure centrally in the psychology of procedural justice. Thibaut and Walker's early work on the subject posits that people view procedures for making distribution decisions as fair to the extent that they are afforded some measure of control over the decision process.<sup>146</sup> This control, which Lind and Tyler refer to as "procedural voice,"<sup>147</sup> powerfully influences the perceived fairness of distributive outcomes, participant satisfaction with outcomes, and attitudes towards decision makers and their institutions.<sup>148</sup>

As previously described, procedural justice researchers have applied these principles to the problem of affirmative action stigma.<sup>149</sup> In the affirmative action context, procedural justice models suggest that candidates have some measure of process control over criteria such as past performance or achievement. Consideration of these criteria gives voice to the candidate. Accordingly, decision processes which utilize them can be termed "high voice" procedures. Consideration of criteria such as gender or race on the other hand, over which the individual candidate has no control, function as "low voice" procedures. Their use erodes perceptions of procedural fairness along the lines originally suggested by Thibaut and Walker and later developed by Tyler and Lind.

As Nacoste has suggested, empirical research revealing a stigmatizing effect of affirmative action preferences can also be understood along these lines.<sup>150</sup> Where merit-related, high voice criteria predominate in decision making, affirmative action programs are perceived as fair, even if group membership is also considered. However, where group

---

142. See generally E. WALSTER ET AL., *EQUITY* (1978).

143. For a review of relevant research, see Tyler et al., *supra* note 66, at 45-52.

144. See Morton Deutsch, *Equity, Equality, & Need: What Determines Which Value Will Be Used as the Basis for Distributive Justice?*, 31 J. Soc. ISSUES 137 (1975).

145. See *id.* See generally, TYLER ET AL., *supra* note 66, at 56-65 (describing research and theoretical models bearing on the choice of allocation decision rules).

146. See John Thibaut & Laurens Walker, *PROCEDURAL JUSTICE* 81-96 (1975).

147. E. ALLEN LIND & TOM R. TYLER, *THE SOCIAL PSYCHOLOGY OF PROCEDURAL JUSTICE* 9 (1988).

148. See *id.* at 180, 191-97; see also TOM R. TYLER, *WHY PEOPLE OBEY THE LAW* 125-34 (1990).

149. See sources cited, *supra* notes 66-71.

150. See Rupert Barnes Nacoste, *Sources of Stigma: Analyzing the Psychology of Affirmative Action*, 12 LAW & POL'Y 175 (1990).

membership is seen as dominating other decision criteria, affirmative action policies are perceived as unfair and are associated with negative reactions to both beneficiaries and to the institutions employing them.<sup>151</sup>

As this research suggests, the problem of merit should be neither trivialized nor ignored. Some conception of merit is essential for any public institution faced with the task of distributing scarce and valued resources. That said, the concept of merit cannot be taken at face value. Those who champion colorblind systems of selection err by presuming that merit is an objective, race- and gender-neutral construct, itself unaffected by subtle forms of intergroup bias.

## 2. *Social Identity, Social Cognition, and the Construction of Merit*

In reality, merit is neither neutral nor objective. It is defined and assessed through the same complex, largely unconscious cognitive processes which subtly bias social judgment in other contexts and give rise to more easily recognizable forms of discrimination.

I have discussed at some length above how schematic expectancies, salience effects, and other forms of cognitive bias can unintentionally influence one's subjective evaluation of members of other social groups. In earlier work, I have attempted to demonstrate that existing civil rights laws, in particular those relating to discrimination in employment, are ill-equipped to reckon with bias of this sort.<sup>152</sup> But in the context of educational admissions, the problem is somewhat different. In education, admissions decisions are often made on the basis of a candidate's performance on anonymously scored, objective assessment measures, such as the Scholastic Aptitude Test (SAT), the Law School Admissions Test (LSAT), or on grade point averages. Grading in many educational settings is conducted on a blind basis, where instructors do not know the identity of the student being evaluated. In these contexts, many of the social judgment biases previously discussed are inoperative.

But it does not follow that selection systems based on these ostensibly objective measures of merit are unaffected by intergroup bias. Such biases intrude in a variety of ways: in the conceptualization of merit, in the selection and use of particular tools for measuring merit, and, more fundamentally, in the very selection of individual merit, however construed and measured, as the applicable allocation decision rule.

Merit-oriented approaches to employment or educational admissions selection call upon decision makers to identify a set of criteria that are expected to predict success in the relevant enterprise. Determining whether a particular criterion accurately predicts success in a

---

151. See *id.* at 183.

152. See Krieger, *supra* note 42.

given endeavor is actually quite a complicated process, governed by a set of complex statistical procedures known as "validation."<sup>153</sup> Although various forms of validation differ in certain ways, they all entail: (1) the identification of behaviors constituting success; (2) the identification of specific claimed predictors of success; and (3) the calculation of the extent and statistical significance, if any, of covariation between these two.

Informal validity judgments, such as whether or to what extent LSAT scores predict success in lawyering, are subject to a variety of errors and biases at each of these three points. Consider the first step described above. In many situations, there is nothing particularly neutral about the way success is defined. The "intuitive" validity tester<sup>154</sup> can identify or prioritize in subjectively biased ways the particular behaviors constituting success in the relevant endeavor. For example, although success in legal academia ostensibly involves performance in three domains—teaching, research, and service—it is generally accepted in many institutions that research is the one that really counts. Rarely are professors denied tenure for mediocre—or even outright poor—teaching. Rarely is service to the school or the profession determinative when tenure decisions are made. Whether plausible justifications exist for this prioritization is not the point. The point is this: defining success is a subjective process subject to subtle forms of intergroup bias with which existing civil rights law has little capacity to reckon.

Consider a second example. Success in many lawyering jobs, such as those with federal regulatory agencies, is generally assessed through evaluation of performance across a variety of domains, including ability to work with and supervise investigators, the quality of litigation outcomes, the ability to manage a caseload, and the quality of written work. However, when hiring or promotion decisions are made, one of these factors, such as the quality of written work, is often weighted more heavily than the other three.

In many situations there may be good reason for weighting certain factors more heavily than others when defining merit. But social identity theory predicts that intergroup bias will play a role as well.

Social identity research conducted under Tajfel's "minimal group paradigm"<sup>155</sup> has shown that we undervalue the product of an outgroup

---

153. For a thorough discussion of statistical validation procedures, and their use in employment discrimination litigation, see *id.* at 1231-36.

154. I borrow here a term coined by social psychologist Lee Ross. See Lee Ross, *The Intuitive Psychologist and His Shortcomings: Distortions in the Attribution Process*, 10 *ADVANCES IN EXPERIMENTAL SOCIAL PSYCHOLOGY* 173 (1977), reprinted in *COGNITIVE THEORIES IN SOCIAL PSYCHOLOGY* 337 (L. Berkowitz ed., 1978).

155. See Tajfel, *supra* note 81. In the minimal group paradigm, subjects are divided into categories on the basis of some arbitrary or "minimal" distinction. For example, experimenters tell

in relation to our own group's products.<sup>156</sup> Accordingly, if those who control the definition of merit in a particular social context belong to the same social reference group, and if members of that group tend to excel on one of the relevant performance dimensions in relation to the others, social identity theory predicts that those who determine merit will tend to overvalue performance in the domain where they collectively excel. Behavioral domains in which a corresponding outgroup excels will tend to be undervalued. In this way, the same ingroup biases that lead to other forms of discrimination influence the definition of merit.

Other cognitive biases influence the particular attributes identified to predict success. The most significant of these biases derives from a phenomenon which Daniel Kahneman and Amos Tversky refer to as the "representativeness" heuristic.<sup>157</sup> According to Kahneman and Tversky, a person attempting to determine whether a particular individual (e.g., a law school applicant) falls into a particular category or class (e.g., successful law school graduates) is unlikely to proceed according to the normative rules of inference governing such decision tasks. Rather, he is likely to match salient attributes of the stimulus object (the applicant) with equally salient attributes characterizing his prototype of the category in question (the successful graduate). If the two sets of attributes match, the person will be more likely to judge the target as a member of the class in question. If the degree of match is low, he will be less likely to reach that conclusion.

This heuristic method of predicting future success injects substantial subjectivity into the assessment of merit. As discussed above, the mental representation of successful law school graduate held by faculty members at prestigious law schools is subject to a variety of subtle biases, including the tendency to overvalue members of one's own group, and the tasks at which one personally excels. Thus, when primed with the phrase "successful law school graduate," legal academics at prestigious universities would be relatively unlikely to form an image of, say, Johnny Cochran, even though his defense in the O.J. Simpson murder trial arguably constitutes one of the most significant advocacy successes in recent criminal litigation history. More ready to mind would be a partner at a large law firm, a successful legal academic, or an appellate

---

subjects that they are being placed in different groups based on having over or under-estimated the size of dots displayed on a screen. After introducing the concept of "groupness," experimenters then investigate the nature and extent of intergroup bias displayed by subjects, as compared to the subjects' behavior in an unassigned control condition. For a review of research conducted under this paradigm, see generally Brewer, *supra* note 80.

156. See Bass & Dunteman, *supra* note 87; Doise, *supra* note 87.

157. See Daniel Kahneman & Amos Tversky, *On the Psychology of Prediction*, 80 PSYCHOL. REV. 237 (1973).

judge. Accordingly, the attributes associated with these prototypes will be subjectively overvalued in merit assessments. The process of selecting a prototype against which applicant attributes come to be matched is profoundly influenced by patterns of intergroup bias, yet this process is almost entirely invisible in antidiscrimination law.

Good lawyers do a wide variety of things and employ a wide variety of skills. Lawyers question and listen to clients, witnesses, decision makers, and opponents. They counsel, comfort, and cajole. They negotiate. They work collaboratively with clients and others to solve problems in ways that require practical wisdom as well as legal analytical skill. Lawyers are public speakers. Lawyers are storytellers. Relatively little of what many lawyers do involves the kind of legal analysis taught in law schools.<sup>158</sup> Which law school applicants will become good lawyers is a question legal educators have barely begun to ask.

Why is it that law schools have made so little attempt to predict applicants' potential as interviewers, counselors, negotiators, creative and flexible problem solvers, storytellers, or public speakers? And what impact does that failure have on the racial, ethnic, or gender distribution of those applicants deemed most qualified for admission to law school? The answer to these questions has much to do with subtle forms of intergroup bias.

Consider the use of LSAT scores in law school admission decision making. The LSAT predicts, within a certain range of error, performance in the first year of law school. However, it does not purport to predict success in lawyering. In the language of employment discrimination law, it has only limited "training program validity," not "job validity."<sup>159</sup> Nonetheless, decision makers use the LSAT extensively in law school admissions, despite its substantial adverse impact on African-American and Latino/Chicano applicants, and despite its limited utility in predicting success in lawyering.

Admittedly, there are good reasons for using LSAT scores in the law school admissions process. It has more demonstrated predictive utility, at least with respect to first year performance, than most available indicators. It is relatively convenient and inexpensive to use. But, if a selection device with such limited validity so negatively affected members of the social group controlling law school admissions procedures, would its use continue? Would its continued use be perceived as fair?

---

158. For a discussion of the sharp divergence between what law students learn and what lawyers do, see, e.g., Bryant G. Garth & Joanne Martin, *Law Schools and the Construction of Competence*, 43 J. LEGAL EDUC. 469 (1993).

159. For cases regarding training program validity, see, e.g., *Washington v. Davis*, 426 U.S. 229 (1976) (Brennan, J., dissenting); *United States v. South Carolina*, 445 F. Supp. 1094 (1977), *aff'd*, 434 U.S. 1026 (1978).

There is substantial reason to suspect not. As an intriguing study by S.H. Ng demonstrates,<sup>160</sup> ingroup favoritism alters people's judgments about the rules of fairness that should be applied in allocating resources or rewards. After being divided into groups according to Tajfel's minimal group technique, subjects in Ng's study observed the performance of an ingroup member and an outgroup member while they worked together on a cooperative task. In one condition, the ingroup member objectively outperformed the outgroup member; in another condition, the situation was reversed. After viewing performance outcomes, subjects were asked to allocate a reward between the two performers and were given a choice of two principles for doing so: an equal allocation principle and an equity principle in which allocations were based on relative performance.

When the outgroup member outperformed the ingroup member, subjects disproportionately applied the equal allocation principle, a choice justified by the cooperative nature of the task. However, when the ingroup member outperformed the outgroup member, subjects disproportionately applied the equity principle and allocated a greater reward to the ingroup member, a result which subjects justified on grounds of relative merit.<sup>161</sup>

Ng's study demonstrates that the fairness rules chosen to govern the allocation of social resources, including those grounded in conceptions of merit, are not preordained manifestations of natural law. They are social, ultimately political decisions, subject to the same forms of intergroup bias that influence reward allocation in other intergroup contexts.

Social identity research on intergroup attribution bias also predicts that the LSAT would be valued differently in making law school admissions decisions if it disproportionally disadvantaged members of the social group controlling those admissions. Once people are divided into social categories, even on a minimal basis, they will tend to attribute success by ingroup members to stable, dispositional traits and failure by ingroup members to external, situational factors. The pattern is reversed with respect to attributions of success and failure by outgroup members.<sup>162</sup> Given these tendencies, we can expect that relative success by ingroup members on a particular task would be attributed to dispositional factors, namely those manifesting qualification or merit. We would further expect that relatively lower levels of success by outgroup

---

160. See Sik Hung Ng, *Equity and Social Categorization Effects on Intergroup Allocation of Rewards*, 23 BRIT. J. SOC. PSYCHOL. 165 (1984). Tajfel's minimal group technique is described *supra*, text accompanying notes 80-85.

161. See *id.* at 170-71.

162. For a review of this literature, see Wilder, *supra* note 84.

members would be attributed to dispositional factors, and would be interpreted as indicating a relative lack of qualification or merit. But if the situation were reversed, and the outgroup outperformed the ingroup, intergroup attribution models predict that relative ingroup failure and relative outgroup success would be disproportionately attributed to external factors, such as the unfair use of an invalid or otherwise biased assessment tool. Here, as elsewhere, subtle forms of intergroup bias affect the definition of merit and the selection of those principles and tools by which it will be measured.

### *3. Intergroup Bias in the Construction of Merit: The Inadequacy of Existing Law*

Preferential forms of affirmative action have largely masked the subtle forms of intergroup bias inherent in the construction of merit. Because preferences could be used to equalize results at the bottom line, the task of identifying such sources of bias and of devising the legal tools to control them have not loomed as pressing priorities. But in a post-affirmative action environment, where the application of previously unquestioned criteria can result in glaring racial and ethnic inequities, the problem assumes a far sharper sense of urgency.

This raises the obvious question of whether, in a post-affirmative action environment, remaining legal tools will prove equal to the task of identifying and eliminating subtle forms of intergroup bias in the construction and assessment of merit. There is ample reason to fear not.

Consider in the first instance the utility of equal protection theory in addressing any of the subtle forms of bias discussed above. Disparate impact theory, which might be used to challenge the use of unvalidated selection criteria, is unavailable in cases brought under the Equal Protection Clause. In such cases, a plaintiff must show that the defendant acted with a discriminatory purpose.<sup>163</sup>

With respect to this requirement, it is not sufficient for the plaintiff to demonstrate that those who chose to use a particular selection criterion were aware that it would disadvantage members of a particular social group. To succeed under an equal protection theory, a plaintiff must prove that the defendant chose the particular selection criterion "because of," not merely "in spite of," its negative impact on members of the plaintiff class.<sup>164</sup> Thus, so far as the establishment of selection criteria is concerned, an entire range of discriminatory practices stemming from subtle forms of intergroup bias is effectively insulated from constitutional challenge.

---

163. See *Washington v. Davis*, 426 U.S. 229, 239 (1976).

164. See *Personnel Adm'r of Massachusetts v. Feeney*, 442 U.S. 256, 279 (1979).



The situation is little better with respect to potential statutory claims. The rule enunciated in *Personnel Administrator v. Feeney* applies in statutory discrimination cases as well. Accordingly, in a disparate treatment case commenced under either Title VI (education) or Title VII (employment) of the Civil Rights Act of 1964, a plaintiff must prove *purposeful* discrimination along the lines *Feeney* describes.<sup>165</sup>

Of course, in employment discrimination cases brought under Title VII, disparate impact theory is potentially available to address certain forms of bias in the construction and assessment of merit. Thus for example, requiring that an objective test which has a disparate impact on a group protected by Title VII be statistically validated in accordance with professional standards<sup>166</sup> prevents many forms of intergroup bias implicated in the ostensibly neutral assessment of merit. Disparate impact theory is an essential equality tool in this context for the simple reason that it renders the definition of merit contestable. Without such a tool, and given the rule enunciated in *Feeney*, there would be no way to challenge the many subtle forms of bias operationalized in the ostensibly group-neutral assessment of merit.

Largely absent from the debate over affirmative action, at least in the educational context, has been any serious recognition that existing legal tools fail to provide a mechanism for contesting unquestioned definitions of merit. In *Regents of the University of California v. Bakke*,<sup>167</sup> a majority of the Supreme Court held that Title VI of the Civil Rights Act of 1964<sup>168</sup> proscribed only those forms of purposeful discrimination prohibited by the Equal Protection Clause.<sup>169</sup> In *Guardians Association v. Civil Service Commission of the City of New York*,<sup>170</sup> a majority of justices, expressing their views in four separate opinions, held again that a violation of Title VI required proof of discriminatory intent.

By contrast, the regulations interpreting Title VI promulgated by the Department of Education's Office of Civil Rights proscribe not

---

165. See *Goodman v. Lukens Steel Co.*, 482 U.S. 656, 684 (1987); *Forsberg v. Pacific Northwestern Tel. Co.*, 840 F.2d 1409, 1418 (9th Cir. 1988); *American Nurses' Ass'n v. Illinois*, 783 F.2d 716, 722 (7th Cir. 1986); *American Fed'n of State, City and Mun. Employees v. Washington*, 770 F.2d 1401, 1405 (9th Cir. 1985).

166. See *Dothard v. Rawlinson*, 433 U.S. 321, 331 (1977); *Albermarle Paper Co. v. Moody*, 422 U.S. 405, 431 (1975) (describing the validation requirement); 29 C.F.R. §1607.1-.7 (1994) (stating that the use of any selection criterion or procedure which has an adverse impact will be considered discriminatory unless it has been validated in accordance with the standards set forth in the Guidelines).

167. 438 U.S. 265 (1978).

168. 42 U.S.C. § 2000d (1998).

169. See *Bakke*, 438 U.S. at 287 (Powell, J.); *id.* at 328 (Brennan, J., joined by White, Marshall, and Blackmun, JJ.).

170. 463 U.S. 582 (1983).

only intentional discrimination, but also the use of criteria or methods of administration that have an unjustified disparate impact on a group protected by Title VI.<sup>171</sup> Oddly, in *Guardians*, a majority of then sitting justices held that even though a plaintiff must prove discriminatory intent to establish a Title VI violation, proof of unjustified discriminatory effects would provide a basis for limited injunctive and declaratory relief in actions brought by private parties under the Title VI regulations.<sup>172</sup>

Although various lower federal courts have followed *Guardians* and permitted Title VI plaintiffs to proceed under a disparate impact theory in actions to enforce the regulations,<sup>173</sup> no reported case has ever challenged the use of either the SAT, the LSAT, the Graduate Record Exam (GRE), or the Medical College Admissions Test (MCAT). Indeed,

171. Specifically, the relevant portion of the regulations provides:

A recipient, in determining the types of services, financial aid, or other benefits, or facilities which will be provided under any such program, or the class of individuals to whom, or the situations in which, such services, financial aid, other benefits, or facilities will be provided under any such program, or the class of individuals to be afforded an opportunity to participate in any such program, may not, directly or through contractual or other arrangements, utilize criteria or methods of administration which have the effect of subjecting individuals to discrimination because of their race, color, or national origin, or have the effect of defeating or substantially impairing accomplishment of the objectives of the program as respect individuals of a particular race, color, or national origin.

34 C.F.R. §100.3(b)(vii)(2) (1997). Department of Education regulations promulgated under Title IX are even more explicit in proscribing the use of tests or other admissions criteria which have a disparate impact based on sex. Pertinent Title IX regulations provide, in relevant part:

A recipient shall not administer or operate any test or other criterion for admission which has a disproportionately adverse effect on persons on the basis of sex unless the use of such test or criterion is shown to predict validly success in the education program or activity in question and alternative tests or criteria which do not have such a disproportionately adverse effect are shown to be unavailable.

34 C.F.R. §106.21(a)(2) (1997).

172. *Guardians*, 463 U.S. at 607 n.27. This principle was reaffirmed a year later in *Alexander v. Choate*, 469 U.S. 287 (1985). Describing its holding in *Guardians*, the *Alexander* Court reflected:

In essence, then, we held that Title VI had delegated to the agencies in this first instance the complex determination of what sorts of disparate impacts upon minorities constituted sufficiently significant social problems, and were readily enough remediable, to warrant altering the practices of the federal grantees that had produced those impacts.

469 U.S. at 293-94.

173. See, e.g., *Villanueva v. Carere*, 85 F.3d 481, 486 (10th Cir. 1996); *New York Urban League, Inc. v. New York*, 71 F.3d 1031, 1036 (2nd Cir. 1995); *Chicago v. Lindley*, 66 F.3d 819, 827 (7th Cir. 1995); *Elston v. Talledega County Bd. of Educ.*, 997 F.2d 1394, 1406 (11th Cir. 1993); *David K. v. Lane*, 839 F.2d 1265, 1274 (7th Cir. 1988); *Gomez v. Illinois State Bd. of Educ.*, 811 F.2d 1030, 1044 (7th Cir. 1987); *Latinos Unidos de Chelsea En Accion (LUCHA) v. Secretary of Housing and Urban Dev.*, 799 F.2d 774, 795 (1st Cir. 1986); *United States v. LULAC*, 793 F.2d 636, 648 (5th Cir. 1986); *Larry P. v. Riles*, 793 F.2d 969, 981, *as amended on denial of reh'g and reh'g en banc* (9th Cir. 1986); *Castaneda v. Pickard*, 781 F.2d 456, 466 (5th Cir. 1986); *Georgia State Conference of Branches of NAACP v. Georgia*, 775 F.2d 1403, 1416 (11th Cir. 1985); *Young v. Montgomery County Bd. of Educ.*, 922 F. Supp. 544 (M.D. Ala. 1996); *Ass'n of Mexican-American Educators v. California*, 836 F. Supp. 1534, 1545 (N.D. Cal. 1993); *Grimes v. Sobol*, 832 F. Supp. 704, 709 (S.D.N.Y. 1993); *Groves v. Alabama State Bd. of Educ.*, 776 F. Supp. 1518, 1522 (M.D. Ala. 1991); *Theresa P. v. Berkeley Unified Sch. Dist.*, 724 F. Supp. 698, 716 (N.D. Cal. 1989).

as of the writing of this Article, I have been unable to find a single reported Title VI or Title IX case in which college or graduate school admissions criteria have been challenged. Thus, unlike employers, whose selection procedures have for years been subject to challenge under Title VII, institutions of higher education have been left to define and assess merit in admissions decision making in an atmosphere utterly devoid of legal contest.

It remains to be seen whether the odd rule of *Guardians* will survive changes in the composition of the Court. Only one of the justices, John Paul Stevens, who held that a disparate impact challenge could be brought under the Title VI regulations, remains on the Court. Chief Justice Rehnquist and Justice O'Connor, both of whom rejected such a rule, also remain. It is difficult to imagine that Justices Scalia, Kennedy, or Thomas would endorse a rule whereby a civil rights enforcement agency would be permitted, through its administrative rule-making power, to impose liability based on an "effects" standard where the Supreme Court has interpreted the relevant statute as requiring proof of intent to discriminate. The dearth of activity under Title VI may, among other things, reflect a lack of confidence in the viability of the *Guardians* rule.

One should not overstate the efficacy of disparate impact theory in the employment context either. As Elizabeth Bartholet has noted, while courts have shown a willingness to require validation of employee selection criteria in cases involving low-level jobs, they have been largely unwilling to require it, or to apply disparate impact theory at all, in cases involving "jobs in high places."<sup>174</sup>

Indeed, even in the employment context, the continued viability of the validation requirement stands in serious doubt. In *Watson v. Fort Worth Bank and Trust*,<sup>175</sup> a plurality of the Court rejected the argument that statistical validation was necessarily required to prove business necessity, and suggested that the existence of a legitimate business reason might suffice to sustain a challenged selection procedure.<sup>176</sup>

Precisely how the courts will interpret the business necessity standard remains highly uncertain, even after passage of the Civil Rights Act of 1991.<sup>177</sup> Thus, there is no way of knowing whether Title VII will

---

174. Elizabeth Bartholet, *Application of Title VII to Jobs in High Places*, 95 HARV. L. REV. 947, 947 (1982).

175. 487 U.S. 977, 999 (1988).

176. See *id.* at 997-99.

177. Largely in response to the Court's decision in the disparate impact case of *Ward's Cove Packing Co., Inc. v. Atonio*, 490 U.S. 642 (1989), Congress enacted the Civil Rights Act of 1991, Pub. L. No. 102-166, 105 Stat. 1071 (1991). Among other things, the Act restored the burdens of proof in disparate impact cases as they existed prior to the Court's decision in *Ward's Cove*. See §§ 2-3, codified at 42 U.S.C. § 2000e-2(k) (1994). However, the Act makes no mention of the validation

continue to provide a business necessity standard stringent enough to control the subtle forms of intergroup bias operationalized in ostensibly objective employee selection procedures.

More fundamentally, however the business necessity standard is construed, disparate impact theory cannot be used to challenge the manner in which an employer prioritizes performance across different job-relevant dimensions. Nor, given the intent requirement, are such subjective valuations amenable to challenge under a disparate treatment theory. Thus, if an employer chooses to value or to reward most highly those attributes or aspects of performance on which white males disproportionately excel, there is little in existing antidiscrimination law that can be mobilized to stop them.

Merit is not a neutral construct, nor is its definition, cultivation, or identification free of intergroup bias. If, as a society, we design equal opportunity policy around the concept of merit but fail to recognize the biases inherent in its construction or application, we will simply operationalize subtle forms of bias in ways that make discrimination extremely hard to identify or correct. Our civil rights laws are not presently equipped to reckon with biases of this sort. In short, basing allocation decisions on group-neutral conceptions of merit is an august goal, but like colorblindness, one more elusive than affirmative action opponents admit.

### C. *The Limits of Individualized Adjudication*

In a 1991 study of federal Title VII litigation, economists John Donohue and Peter Siegelman observed that over the course of the 1980s, the federal employment discrimination docket changed in a variety of significant ways.<sup>178</sup> Specifically, Donohue and Siegelman showed that while early Title VII cases tended to challenge discrimination in hiring, by the end of the 1980s most cases alleged discrimination in discharge.<sup>179</sup> The 1980s also witnessed a dramatic decline in the use of the class action device in employment discrimination litigation, such that by the end of the decade the overwhelming majority of Title VII suits involved individual claims of disparate treatment discrimination brought by individual private litigants.<sup>180</sup> Claims of disparate impact,

---

requirement, and leaves unresolved the precise contours of the business justification defense. See 42 U.S.C. § 2000e-2(k)(1)(A) (1992).

178. See John J. Donohue III & Peter Siegelman, *The Changing Nature of Employment Discrimination Litigation*, 43 STAN. L. REV. 983 (1991).

179. See *id.* at 984.

180. See *id.* at 1019 (describing the decline in the incidence of employment discrimination class actions and attributing that decline in substantial part to changes in legal rules relating to class certification standards under Rule 23, as well as standards for awards of attorneys' fees).

which are almost by nature class based, constituted less than two percent of the courts' Title VII caseload during that decade.<sup>181</sup>

Although some might argue that a decline in hiring discrimination cases signals the disappearance of discrimination from the labor market, Donohue and Siegelman suggest<sup>182</sup> and recent investigations indicate that this is not the case. For example, a study by the Urban Institute involving black and white testers in Washington, D.C. and Chicago, Illinois provides substantial evidence that discrimination against black applicants for employment is still widespread.<sup>183</sup> A similar study conducted by the Fair Employment Council of Greater Washington in 1990 and 1991 revealed that black testers received outcomes less favorable than white testers 24% of the time, and that Hispanic testers were treated less favorably than white testers 22% of the time.<sup>184</sup> In those cases in which job offers were given to both white and black testers, employers offered whites a higher initial wage 16.7% of the time.<sup>185</sup> Yet a third tester study conducted by the General Accounting Office found that Hispanic testers were three times more likely to receive unfavorable treatment, receiving 25% fewer interviews and 34% fewer job offers than white testers with the same qualifications.<sup>186</sup>

The effect is not confined to race and ethnicity. In a 1995 study, researcher David Neumark and his collaborators sent resumes of male and female applicants with the same qualifications to restaurants in Philadelphia. In the high-priced restaurants, male applicants were more than twice as likely to be called for an interview and five times more likely to be offered employment than female applicants.<sup>187</sup>

Discrimination in employment persists in the American labor market. Changes in substantive and procedural rules, along with changes in the standards governing awards of attorneys' fees to prevailing plaintiffs, have reduced the effectiveness of the class action and the disparate impact case as tools for remedying discrimination. Individual disparate treatment litigation is now the primary mechanism for redressing discrimination in employment. If affirmative action programs are eliminated, and with them incentives for ensuring that women and minorities

---

181. See *id.* at 989, 998.

182. See *id.* at 1015, 1019.

183. See MARGERY A. TURNER ET AL., OPPORTUNITIES DENIED, OPPORTUNITIES DIMINISHED: RACIAL DISCRIMINATION IN HIRING (showing white testers 16% more likely to receive job offers than black testers).

184. See Marc Bendick, Jr. et al., *Measuring Employment Discrimination Through Controlled Experiments*, 23 REV. BLACK POL. ECON. 25, 29 (1994).

185. See *id.* at 32.

186. See United States General Accounting Office, *Immigration Reform: Employer Sanctions and the Question of Discrimination*, GAO/GDD-90-62 (1990).

187. See DAVID NEUMARK ET AL., SEX DISCRIMINATION IN RESTAURANT HIRING: AN AUDIT STUDY 7-10 (National Bureau of Econ. Research Working Paper No. 5024, 1995).

remain proportionately represented in a particular job category or educational program, the search for individual instances of discrimination will assume even greater significance in the implementation of the nation's equal opportunity policy. In essence, enforcement of the nation's stated commitment to equal employment opportunity has already come to depend, and absent affirmative action programs will even more substantially depend, on the ability of individual victims to recognize that they have been discriminated against, and on the ability of juries and other fact finders<sup>188</sup> accurately to discern whether discrimination has occurred.

The individual disparate treatment case is an extremely weak enforcement tool, poorly calculated to do the task assigned it in a post-affirmative action environment. Many of the arguments supporting this view have been developed elsewhere and need not be repeated here.<sup>189</sup> My present purpose is to show, from a social and cognitive psychological perspective, that reliance on the individual disparate treatment adjudications will result in the significant underidentification of discrimination, not only by decision makers, as I have argued previously,<sup>190</sup> but by victims and fact finders as well.

The basis for this claim can be summarized as follows. First, as a simple consequence of cognitive constraint, people are relatively poor at identifying discrimination on a case-by-case as opposed to aggregate basis. This difficulty is exacerbated when discrimination is "non-prototypical:" when it fails to reflect people's expectations as to what discrimination looks like.

Furthermore, people have difficulty noticing things that do not happen in comparison to things that do. Thus, the subtle forms of ingroup preference which constitute what social cognition theorists refer to as "Type II discrimination,"<sup>191</sup> can be expected to go largely unnoticed by both victims and decision makers. Over time, these subtle,

---

188. Throughout this Section, I use the term "fact finders" in the broad sense—to refer not only to judges in bench trials, but also to human resources professionals, government equal opportunity administrators, alternative dispute resolution providers, and legal advisers responsible for reviewing challenged decisions and deciding whether to take or recommend voluntary remedial action.

189. See, e.g., Krieger, *supra* note 42, at 1238-39 (predicting a high false negative error rate in disparate treatment adjudications under current legal standards); David A. Strauss, *The Law and Economics of Racial Discrimination in Employment: The Case for Numerical Standards*, 79 GEO. L. J. 1619 (1991) (arguing from an economic efficiency standpoint that objectives of the antidiscrimination laws are not best served by trying to detect individual acts of discrimination and suggesting that the error rate in disparate treatment adjudication is likely to be high). See also Thomas A. Cunliff, *The Price of Equal Opportunity: The Efficiency of Title VII After Hicks*, 45 CASE W. RES. L. REV. 507 (1995) (arguing that the Court's decision in *St. Mary's Honor Ctr. v. Hicks*, 509 U.S. 502 (1993), has decreased the efficiency of Title VII by increasing the probability of false negative error in disparate treatment adjudications).

190. See Krieger, *supra* note 42, at 1213-15.

191. See *infra* text accompanying notes 226-227.

incremental forms of advantage result in the creation of objective disparities between the qualifications of preferred ingroup versus non-preferred outgroup members. By the time a negative action is taken, noticed by the victim, and made the subject of a discrimination claim, the claim frequently can be successfully defended on the grounds that the favored employee or applicant was "objectively more qualified." Thus fact finders, as well as victims and decision makers, are unlikely to recognize discrimination caused by Type II bias.

Finally, determining whether discrimination has occurred is largely an exercise in causal attribution. Unfortunately, the same biases in causal attribution that commonly distort intergroup perception and judgment and result in unintentional forms of discrimination can be expected to distort fact-finder determinations as to whether discrimination has occurred in any particular case. We cannot expect the strength or direction of intergroup *decision-maker* bias to vary significantly from the strength or direction of intergroup *fact-finder* bias. To the contrary, the level of decision-maker bias in a given judicial district will likely be mirrored by the level of fact-finder bias in that district. Disparate treatment adjudication therefore can be expected to pack little "socio-therapeutic" power in either identifying or remedying subtle forms of intergroup bias. In fact, the conceptual structure of disparate treatment analysis can be expected to potentiate stereotypes, thus exacerbating fact-finder bias and increasing the incidence of false negative error in formal or informal<sup>192</sup> disparate treatment adjudications.

### 1. *Individualized Adjudication and the Problem of Case-By-Case Data*

In a 1986 experiment, psychologist Faye Crosby and her colleagues at Yale University<sup>193</sup> sought to explore the effect of format on subjects' abilities to discern patterns of discrimination in data reflecting comparative levels of compensation accorded male and female managers at a hypothetical company. Male subjects were provided with a set of materials about "Company Z," comprised of ten different departments. The materials were constructed so that the women in the company earned, on the average, significantly less than the men, controlling for various input factors such as position, seniority, education, and supervisory rating. In one condition, subjects first received the relevant information in aggregated form—data reflecting salary and performance on the input variables for a typical man and a typical woman from each of

---

192. By "informal" disparate treatment adjudications I refer to management or government administrative review of contested decisions, internal grievance procedures, and various forms of alternative dispute resolution.

193. See Faye Crosby et al., *Cognitive Biases in the Perception of Discrimination: The Importance of Format*, 14 SEX ROLES 637 (1986).

the ten departments was systematically arranged on a single large page. After viewing this material, subjects were asked whether sex discrimination existed in the company as a whole.<sup>194</sup> After seeing the information in aggregate form, subjects were then shown the same data on a department by department basis, and were asked to make serial judgments about each department.

In a second condition, researchers reversed the order of these two steps. Thus subjects were first asked to make serial department by department judgments, based on disaggregated data about each department. They were then shown the aggregated data and asked to determine whether sex discrimination existed in the company as a whole.

Crosby and her colleagues learned that format had a significant effect on subjects' abilities to identify sex-linked disparities embedded in the stimulus materials. When information was presented on a department by department basis, subjects perceived significantly less discrimination than when the same information was presented in aggregated form. Whether the aggregated data or the disaggregated data was presented first made little difference.<sup>195</sup> In both conditions, subjects more easily identified discrimination when information was presented on a company-wide basis than on a department by department basis.

One could criticize the Crosby study on the grounds that in the aggregated format, subjects were asked to make determinations regarding discrimination in the company as a whole, whereas in the disaggregated format they were asked to make determinations about discrimination in one particular department, as to which they had much less information. Thus, the results obtained could have simply reflected subjects' quite rational unwillingness to draw conclusions based on a small data set rather than on any effect of format, *per se*.

To correct this problem, Stanford researcher Diane Cordova replicated Crosby's 1986 study with a modified experimental design.<sup>196</sup> Cordova's subjects reviewed data in one of three different formats: aggregate, case-by-case/immediate judgment, and case-by-case/delayed judgment. The aggregate condition was the same as in the original

---

194. Based on criticisms that use of the emotionally-charged term "sex discrimination" might have influenced the result, another research term was substituted. Instead of being asked to judge whether sex discrimination existed, subjects were asked to discern differences in the patterns of compensation for managers at a hypothetical Plant A and Plant B. Catherine Twiss et al., *Affirmative Action and Aggregate Data: The Importance of Patterns in the Perception of Discrimination*, in *AFFIRMATIVE ACTION IN PERSPECTIVE* 159 (Fletcher A. Blanchard & Faye J. Crosby eds., 1989). The results obtained by Twiss and her colleagues replicated those in Crosby's original 1986 study.

195. Nor did self-reported political ideology significantly effect the results. Subjects labeling themselves as feminist and subjects labeling themselves as "anti-feminist" were equally susceptible to the effects of information format. *See id.* at 643.

196. *See* Diane I. Cordova, *Cognitive Limitations and Affirmative Action: The Effects of Aggregate Versus Sequential Data in the Perception of Discrimination*, 5 Soc. JUST. RES. 319 (1992).



Crosby study: relevant input factor and salary information for all ten departments was arranged for subjects' review on one large page. In the case-by-case/immediate judgment condition, subjects were presented with the same information one department at a time, with the order of the departments randomized to control for primacy effects. After the information for each department was presented, subjects were asked to state whether that information indicated the existence of sex discrimination in that particular department. In the case-by-case/delayed judgment condition, subjects were presented with the relevant information on the same department by department basis, but reviewed such information for each of the ten departments before proffering a single judgment as to discrimination in the company as a whole.<sup>197</sup>

The results were striking. Subjects, whether male or female, more readily identified sex discrimination in the aggregate condition than in either the case-by-case/immediate judgment condition or, more importantly, in the case-by-case/delayed judgment condition.<sup>198</sup> Perhaps even more striking, although subjects in the aggregated and case-by-case conditions were similarly able to recall both the discrepancies in male versus female salary levels and the substantial equivalencies in male and female performance, subjects in the aggregate condition were significantly more willing to *label* the discrepancies in salary as sex discrimination than were the other subjects.<sup>199</sup>

The implications of these findings are readily apparent. Unless and until they file suit, individual victims of discrimination rarely have access to comparison information about even *one* other applicant or employee, let alone aggregated data for a department or company as a whole. Fact finders, like victims, are also unlikely to have access to this kind of information. It is expensive, and in light of its potential disclosure and hostile use in future discrimination litigation, risky to compile. While equal employment opportunity/affirmative action departments once routinely collected such information, and used it to scrutinize hiring, promotion, and compensation decisions for AA/EEO compliance, such departments and standards have largely disappeared from the corporate landscape.<sup>200</sup> Human resources professionals and government bureaucrats reviewing challenged decisions are frequently presented only with information concerning "the case at hand," that is, information comparing characteristics of a particular complainant with characteristics of a single successful applicant or employee. In discharge situations, these decision makers frequently render judgments without reviewing

---

197. See *id.* at 326.

198. See *id.* at 328.

199. See *id.* at 329.

200. See generally THOMAS BYRNE EDSALL & MARY D. EDSALL, CHAIN REACTION (1991).

any comparison information at all, let alone systematically organized, aggregate data reflecting on company-wide patterns.

Like subjects in the format studies, these informal fact finders are unlikely to see discrimination in such case-by-case information. The reasons for this are obvious. When presented in anecdotal form, differences in treatment are too easily attributed to observed differences in one or another input variable.

For example, individual employment discrimination plaintiffs often fail in their attempts to convince courts that the more favorable treatment of an employee of a different race or gender evidences intergroup bias. In many of these "comparability" cases, courts point to observed differences in the precise circumstances surrounding the two incidents and hence find that the employees compared were not "similarly situated."<sup>201</sup> Because two situations are rarely identical in all respects, and because differences in treatment are easily attributed to these differences rather than to discrimination, it is easy to miss the role played by intergroup bias when incidents are considered in isolation. Only when aggregated data are presented in a systematic way can one see these differences cancel each other out, revealing underlying patterns of disparity in the treatment of members of different groups.

Aggregated information, like that presented in the Crosby, Twiss, and Cordova studies, may even be unavailable to fact finders in more formalized individual disparate treatment adjudications. Internal or alternative dispute resolutions procedures rarely provide complainants with access to the breadth of information required to prepare systematically organized aggregated data. This situation may persist even when cases go to court. In individual disparate treatment litigation, defendants frequently, and often successfully, resist discovery targeted at departments or units other than that in which the plaintiff was employed on the theory that only decisions made by the particular supervisor whose action plaintiff is challenging is relevant.<sup>202</sup> Thus, even those victims of

---

201. See *Mitchell v. Toledo Hospital*, 964 F.2d 577, 583-84 (1992) (holding that a black employee who was fired after hiding a file from a supervisor as a "practical joke" was not similarly situated with white employees who were not fired for insubordination or absenteeism); *Moore v. Charlotte*, 754 F.2d 1100, 1107 (4th Cir. 1985) (holding that black officer demoted for fixing a traffic ticket in exchange for money not comparable to white officers who were not demoted after fixing traffic tickets for friends); *Talley v. U.S. Postal Serv.*, 720 F.2d 505 (8th Cir. 1983) (holding that mail carrier who lost keys on two occasions is not comparable to others who lost keys on one occasion); *Meyer v. California & Hawaiian Sugar Co.*, 662 F.2d 637 (9th Cir. 1981) (holding female employee who wrote memo containing racist comments, which fell into hands of black employees and made them angry, not comparable to male employee who made racist comments that did not stir anger in black employees).

202. See, e.g., *Jones v. Frank*, 973 F.2d 673, 676 (8th Cir. 1992); *Scales v. J.C. Bradford & Co.*, 925 F.2d 901, 906-07 (6th Cir. 1991); *Jones v. Gerwens*, 874 F.2d 1534, 1541 (11th Cir. 1989); *Cooper v. North Olmsted*, 795 F.2d 1265, 1271 (6th Cir. 1986); *Prouty v. National R.R. Passenger Corp.*, 99

discrimination who overcome their own cognitive limitations in perceiving discrimination, find a lawyer to represent them, and file suit, may be unable to present information in a manner calculated to afford them success. If, in a post-affirmative action environment, the individual disparate treatment case becomes our primary equal opportunity policy tool, and if victims, jurors, and other fact finders have access only to anecdotal or case-by-case information, we can reasonably predict a significant under-identification of discrimination which no other policy tool will be available to address.

## 2. *Individual Adjudication and the Problem of Schematic Expectancies*

As we have seen, people have a hard time recognizing discrimination on a case-by-case basis. They have an even harder time recognizing discrimination when it is not prototypical, that is, when it does not fit preexisting ideas about what discrimination looks like. In this regard, discrimination is not particularly different than other social phenomena that may become the subject of legal adjudication. As Albert Moore suggests, jurors in a variety of subject matter contexts decide cases by matching the evidence presented at trial with elements of their schematic expectancies or prototypes of cases of the sort at issue.<sup>203</sup> To the extent that the story presented by one or the other party at trial corresponds with jurors' internal, schematic frames of reference, that party's case will seem plausible and persuasive and will be more likely to succeed.

Most people have preconceived notions about what discrimination looks like. These "discrimination schemas" generally include a number of standard features. First, people assume that discrimination is intentional—that the discriminating decision maker *knows* that he is taking the target's group status into account in making the decision, and *means* to do so. Evidence of this common schematic expectancy is well-illustrated in the lead opinion in *Price Waterhouse v. Hopkins*,<sup>204</sup> in which Justice Brennan wrote:

In saying that gender played a motivating part in an employment decision, we mean that, if we asked the employer at the moment of the decision what its reasons were and if we received a truthful response, one of those reasons would be that the applicant or employee was a woman.<sup>205</sup>

---

F.R.D. 545, 548 (D.C. Cir. 1983); *James v. Newspaper Agency Corp.*, 591 F.2d 579, 582 (10th Cir. 1979); *Suggs v. Capital Cities/ABC, Inc.*, 122 F.R.D. 430, 431-32 (S.D.N.Y. 1988).

203. See Albert J. Moore, *Trial by Schema: Cognitive Filters in the Courtroom*, 37 UCLA L. Rev. 273, 276 (1989).

204. 490 U.S. 228 (1989).

205. *Id.* at 250.

Decision-maker consistency is a second standard feature. If a person discriminates against members of a particular gender, racial, or ethnic group, we expect him to do so *consistently*. We assume a particular decision maker is unlikely to have a "taste for discrimination"<sup>206</sup> one day and not the next. For this reason, many courts have adopted what is known as the same actor doctrine,<sup>207</sup> holding that if the same person who hired an employee makes the decision to fire him, a strong inference or presumption of non-discrimination arises.<sup>208</sup> This inference is seen as justified because "it hardly makes sense to hire workers from a group one dislikes (thereby incurring the psychological costs of associating with them), only to fire them once they are on the job."<sup>209</sup>

Indeed, if discrimination is seen as resulting from a "taste" or stable preference, it is hard to understand why a person might be hired in spite of the preference, only to be fired later to satisfy it. Furthermore, if one views discrimination as resulting from a stable *ex ante* preference, one will expect all members of a disfavored social group to be accorded similar negative treatment. Because people assume that discrimination will operate in this way, defendants are often successful in using evidence of fair treatment of members of the claimant's group to prove that discrimination did not occur.<sup>210</sup>

---

206. Economists use the term "taste for discrimination" to describe an aversion to workers of a particular race, sex, national origin, or religion under conditions in which all groups of workers are equally productive. See, e.g., GARY S. BECKER, *THE ECONOMICS OF DISCRIMINATION*, 14, 39-55 (1957); John J. Donahue III, *Is Title VII Efficient?*, 134 U. PA. L. REV. 1411, 1415-20 (1986).

207. Regarding the same actor doctrine, see generally, Marlinee C. Clark, *Discrimination Claims and "Same Actor" Facts: Inference or Evidence?*, U. MEM. L. REV. 183 (1997).

208. Many circuit courts have adopted the same actor doctrine in Title VII cases alleging disparate treatment in termination. See, e.g., *Grady v. Affiliated Century, Inc.*, 130 F.3d 553, 560 (2d Cir. 1997); *Bradley v. Harcourt, Brace & Co.*, 104 F.3d 267, 270-71 (9th Cir. 1996); *Brown v. CSC Logic, Inc.* 82 F.3d 651, 658 (5th Cir. 1996); *Evans v. Technologies Applications & Serv. Co.*, 80 F.3d 954, 959 (4th Cir. 1996); *E.E.O.C. v. Our Lady of Resurrection Med. Ctr.*, 77 F.3d 145, 152 (7th Cir. 1996); *Burhmaster v. Overnite Transp. Co.*, 61 F.3d 461, 463 (6th Cir. 1995), *cert. denied*, 516 U.S. 1078 (1996); *Rand v. CF Indus., Inc.*, 42 F.3d 1139, 1147 (7th Cir. 1994); *LeBlanc v. Great Am. Ins. Co.*, 6 F.3d 836, 847 (1st Cir. 1993); *Lowe v. J. B. Hunt Transp., Inc.*, 963 F.2d 173, 175 (8th Cir. 1992). See also, *Williams v. Vitro Serv. Corp.*, 144 F.3d 1438, 1443 (11th Cir. 1998) (rejecting same actor doctrine as a presumption, but permitting it as an inference). As this Article goes to press, only the Third Circuit has rejected the same actor doctrine. See *Waldron v. SL Indus., Inc.* 56 F.3d 491, 496, n. 6 (3rd Cir. 1995) (holding that same actor facts are simply evidence, like any other, and should not be accorded any special presumptive value.)

209. *Proud v. Stone*, 945 F.2d 796, 797 (4th Cir. 1991), citing, John J. Donohue & Peter B. Siegelman, *The Changing Nature of Employment Discrimination Litigation*, 43 STAN. L. REV. 983, 1017 (1991).

210. See, e.g., *Aramburu v. Boeing Co.*, 112 F.3d 1398, 1406 (10th Cir. 1997) (finding unfavorable treatment of plaintiff insufficient to support an inference of discriminatory animus when other minorities were accorded the same treatment as non-minority employees); *Jefferies v. Harris County Community Action Ass'n.*, 615 F.2d 1025, 1030 (5th Cir. 1980) (holding that plaintiff failed to prove race discrimination based on evidence that an employee of the same race had been promoted); see also *Brazer v. St. Regis Paper Co.*, 498 F. Supp. 1092, 1098 (M.D. Fla. 1980) (finding that

Third, we generally think of discrimination as stemming from hostility or negative animus. For this reason, we speak of discrimination *against* members of other groups, rather than, for example, discrimination *in favor of* members of our own group.<sup>211</sup> Given that we think of discrimination as reflecting a kind of hostility, we expect it to be expressed in objectively unfair treatment of target group members and to be accompanied by racist or sexist comments, or other direct evidence of negative animus. The prototypical discriminator is either a sexist, an ageist, or some other kind of bigot, and he treats members of the disfavored groups objectively unfairly.

Finally, the prevailing discrimination schema reflects an assumption that invidious animus trumps other factors that would, absent discrimination, be used to make the contested decision. This belief in the independence of discrimination explains the view, embedded in disparate treatment jurisprudence, that where racial or gender bias exists, it commonly operates as the sole motive for a discriminatory employment decision. According to this view, intergroup bias supplants rather than distorts a decision maker's evaluation of legitimate input variables. Perhaps for this reason, "mixed motive" cases are conceived in Title VII disparate treatment jurisprudence as the exception and "sole motive" or "pretext" cases the rule.<sup>212</sup>

The standard discrimination schema might be encapsulated as follows: discrimination occurs when a sexist, racist, or otherwise bigoted person makes decisions about members of a targeted group. Because of his negative feelings towards or beliefs about members of the disfavored group, the discriminator purposefully treats them unfairly. His negative feelings are likely to be expressed in racist or sexist comments. Even if the discriminator does not express these negative feelings, he knows of them and would admit them were he being honest. The kind of prejudice that leads to discrimination functions like a personality trait: it is something that exists inside the discriminator. It is relatively stable and expresses itself consistently over time and across different situations.

I have attempted elsewhere to discredit the assumptions of intentionality and independence embedded in the prevailing discrimination schema.<sup>213</sup> I seek here to question two other assumptions: the

---

replacement of plaintiff with an employee of the same race undermined his claim of race discrimination in job assignment).

211. This implicit assumption, that intergroup bias manifests as negative action taken against outgroup members, is reflected in the language of Title VII itself. The statute makes it unlawful for an employer "to *fail or refuse* to hire or to discharge... or otherwise discriminate *against* any individual" because of his race, color, religion, sex, or national origin. Title VII of the Civil Rights Act of 1964, *as amended* by 42 U.S.C. § 2000e-2(a)(1) (1994) (emphasis added).

212. For a thorough discussion of this "sole motive fallacy," and its effect on disparate treatment theory, see Krieger, *supra* note 42, at 1178-79, 1223-24.

213. See *id.* at 1168-77, 1181-85.

expectation of consistency and the expectation of unfair treatment of outgroup members.

*a. Schematic Expectancies and the Problem of Complexity*

As we have seen, many people—among them federal district and circuit court judges and law and economics scholars—seem to believe that discrimination derives from a kind of taste, or stable preference, that expresses itself consistently over time and across varying situations.<sup>214</sup> In this regard, intergroup bias is seen as functioning much like a personality trait.

Even if discrimination derived primarily from such a taste or preference, any model based on this assumption and predicting consistent decision making would be seriously flawed. Empirical social psychology has to a great extent discredited the assumption that people behave consistently over time and across different situations in conformity with stable preferences and personality traits.<sup>215</sup> Even those researchers who support a relatively “strong” theory of personality acknowledge that the relationship between traits and behaviors depends significantly on the situations in which behaviors occur.<sup>216</sup>

Let us assume that at least a substantial proportion of discriminatory decisions derive, as law and economics scholars suppose, from tastes or preferences. Let us further assume, as empirical social and personality psychology has established, that the relationship between such tastes or preferences on the one hand and behavior on the other depends significantly on the situation in which the actor finds himself. To these, let us add the further assumption that a second, substantial set of discriminatory decisions derive not from tastes or preferences, but from the operation of subtle, largely cognitive forms of intergroup bias, and from the operation of judgmental heuristics such as the salience and polarized evaluation effects described in Part I above. The expression of these biases, like the relationship between personality traits and behaviors, are situation-dependent.

If we accept the proposition that the expression of tastes and preferences is situationally determined, and if we accept the proposition that much discrimination derives not from tastes or preferences at all

---

214. See discussion *supra*, in text accompanying notes 206-213.

215. For a comprehensive overview of social psychological research suggesting the weakness of global trait constructs as reliable predictors of behavior and choice, see generally LEE ROSS & RICHARD E. NISBETT, *THE PERSON AND THE SITUATION* (1991).

216. See, e.g., Walter Mischel & Yuichi Shoda, *A Cognitive-Affective System Theory of Personality: Reconceptualizing Situations, Dispositions, Dynamics, and Invariance in Personality Structure*, 102 *PSYCHOL. REV.* 246 (1995); Yuichi Shoda et al., *Links Between Personality Judgments and Contextualized Behavior Patterns: Situation-Behavior Profiles of Personality Prototypes*, 11 *Soc. COGNITION* 399 (1993).

but from the operation of various situation sensitive cognitive biases and heuristics, a startling conclusion follows. Cognitive processes of perception and judgment, both those that lead to judgment error directly and those that moderate the expression of traits or preferences, are not characterized by the simple, almost hydraulic patterns of cause and effect embedded in the prevailing discrimination schema. Rather, they are characterized by complexity.

By this I mean that many different heuristics, biases, and other judgmental influences, such as primacy, recency, or contrast effects, not to mention motivational influences, will bear simultaneously on any given social judgment task. Accordingly, discrimination caused by situation-sensitive or cognitive forms of intergroup bias will not operate consistently. Quite the contrary, the operation and effects of these forms of bias will vary considerably from context to context, as is characteristic of complex systems. Therefore, we can predict that intergroup bias will cause discrimination. But we cannot say that intergroup bias will *always* cause discrimination to occur, nor can we predict exactly *when* discrimination will occur. We might not even be able to identify when discrimination *has* occurred.

In short, we cannot expect systems in which cognitive or other situation-sensitive forms of bias are operating to present neat, consistent patterns of outgroup subordination. The resulting disparities may be less conspicuous. They may be expressed as disparities of magnitude rather than disparities of kind. They may not be statistically significant. They may not present a simple story, easy to tell and consistent with fact finders' schematic expectancies.

Although the implications of this insight could easily expand into a separate article, two points illustrating its significance will serve our present purposes. The first point is this: if people believe that discrimination results from *ex ante* preferences which manifest consistently over time and across different situations, people will have a hard time understanding, for example, how a biased decision maker would make both the decision to hire and the decision to fire a protected class member. Second, if people expect discrimination to be expressed as consistent patterns of outgroup subordination rather than as a set of tendencies which, over time, result in inconsistent but nonetheless substantial disparities in distribution, much discrimination will go unrecognized, and for that reason, unredressed. Let us look a little more closely at each of these problems, and consider their implications for post-affirmative action equal opportunity policy.

i. *Of Single Actors, Ex Ante Preferences, and the Situation-Dependent Nature of Biased Decision Making*

Consider again the same actor doctrine,<sup>217</sup> which provides that where the same person is responsible for both hiring and firing an employee, an inference or presumption should arise that there was no discrimination in the decision to fire. The implicit theory about what discrimination is and how it operates reflected in this doctrine is clearly reflected in the Seventh Circuit's opinion in *E.E.O.C. v. Resurrection Medical Center*.<sup>218</sup> In *Resurrection Medical Center*, the court observed, "If Boettcher [the decision maker] wanted to discriminate against Braddy [the discharged employee] because of her race, she could have refused to hire her in the first place."<sup>219</sup> This observation reflects an important assumption—that discrimination in hiring and discrimination in discharge are caused by the same thing, namely a stable taste or preference for social distance. According to this view, intergroup biases—whether borne of stereotypes or of outright antipathy—serve as *ex ante* decision rules, equally applicable to hiring or discharge decisions. If the *ex ante* rule was not applied in the hiring context, it is hard to understand why it would be applied in a decision to discharge. Within its own assumptive universe, the single actor inference makes good sense. But there is substantial reason to believe that its underlying assumptions are wrong, that this is simply not how discrimination works, at least not discrimination deriving from implicit stereotypes about the aptitudes and abilities associated with different demographic groups.

In this regard, consider a well-known study by Princeton psychologists John Darley and Paget Gross investigating the nature and function of a phenomenon known as the "cognitive confirmation effect."<sup>220</sup> This effect occurs when people with different expectancies are exposed to the same information about a target person. Even though they review the same information or view the same series of events, perceivers with opposite preconceptions tend to emerge with their expectancies confirmed. Darley and Gross sought to identify the specific cognitive processes that mediate this phenomenon.

In their study, Darley and Gross had subjects view one or two videotapes, each portraying a school age child named "Hannah." By manipulating the content of the first videotape, Darley and Gross led subjects to believe that Hannah's family was either disadvantaged or affluent. So, in the "low income" condition, the video depicted Hannah

---

217. See *supra* discussion accompanying notes 207-209.

218. 77 F.3d 145 (7th Cir. 1996).

219. *Id.* at 152.

220. See John M. Darley & Paget H. Gross, *A Hypothesis-Confirming Bias in Labeling Effects*, 44 J. PERSONALITY & SOC. PSYCHOL. 20 (1983).



playing in a stark, fenced-in, asphalt school yard in a run-down urban neighborhood. Researchers told subjects in this condition that Hannah's parents had only a high school education, that her father worked as a meat packer, and that her mother was a seamstress who worked at home.

In the "high income" condition, subjects watched Hannah play in a tree-lined park in an upper middle class neighborhood. The video depicted her school as a large, attractive modern structure, with adjacent playing fields and a shaded playground. It described her father as an attorney, her mother as a free-lance writer.

Immediately after subjects watched one or the other of these videotapes, a subgroup of subjects in each income condition was asked to predict whether Hannah would perform academically "at grade level," "above grade level," or "below grade level." Reluctantly, and often under protest over the paucity of diagnostic information they had received, subjects in both conditions tended to predict that Hannah would perform "at grade level." In other words, they refused to form, or at least express, an initial prediction based on the stereotypes about academic achievement associated with high and low socioeconomic status.

Rather than immediately predicting Hannah's ability level, a second subgroup in each income condition watched a second videotape. This second videotape, which was identical across both conditions, depicted Hannah responding verbally to twenty five achievement test problems. Researchers told subjects that the test included easy, moderate, and difficult problems. Researchers manipulated Hannah's performance in this second video so as to be ambiguous.

After watching this second video, subjects in the low-income and high-income conditions were asked to evaluate Hannah's abilities. At this point, dramatic differences emerged. Subjects in the "affluent Hannah" condition rated her academic ability significantly higher than did subjects in the "disadvantaged Hannah" condition. Open-ended comments revealed differences in perception as well. For example, subjects frequently reported the low-income Hannah as having "difficulty accepting new information," while they frequently described high-income Hannah as demonstrating the "ability to apply what she knows to unfamiliar problems."<sup>221</sup>

Darley and Gross interpreted their findings as suggesting that exposure to mixed or ambiguous diagnostic information actually potentiates social stereotypes. People do not believe it appropriate to base social predictions or judgments on stereotypes. Accordingly, when asked to make a prediction based on socio-economic information alone, subjects

---

221. *Id.* at 28.

protested and kept their predictions close to a neutral mean. Nonetheless, these same stereotypes eschewed by subjects in the immediate response condition found expression in social evaluation tasks performed in a more "information rich" context. How?

Darley and Gross answered this question by suggesting that stereotypes function not as consistent *ex ante* decision rules, but as dormant expectancies, which become activated when people hear about or observe the actions of a stereotyped other. Through the mediation of various cognitive processes, including selective attention, biased interpretation and attribution, and differential recall, a person's judgment is more powerfully influenced by expectancy-consistent information than by expectancy-inconsistent information. In this way, two people can view precisely the same sequence of events and reach radically different conclusions, depending on their prior expectancies.

The implications of these findings for disparate treatment theory in general, and for the same actor inference in particular, should be obvious. Not all discrimination results from the application of tastes or preferences, expressing themselves as consistent, *ex ante* decision rules. Discrimination sometimes results from the operation of implicit stereotypes which bias the interpretation, retention, and utilization of incoming mixed or ambiguous information. The same decision maker who, when individuating information is sparse, refuses to make an initial negative prediction based on stereotypes, may, like the subjects in Darley and Gross's study, fall prey to their biasing effect as more ostensibly diagnostic information becomes available. Cognitive forms of intergroup bias will not operate consistently, even in the same decision maker. Their expression will vary, according to the specific situation in which the decision maker finds himself. A fact finder looking for consistent patterns of discrimination is apt to see no discrimination at all.

*ii. Expectations of Consistency and the Search for Ordinal Inequity*

As we have seen, people have difficulty recognizing discrimination when it is not conspicuous and consistent. In a dramatic illustration of this effect, social psychologist Susan Clayton demonstrated that subjects were unable to recognize patterns of sex discrimination in compensation when the differences in salary were small or inconsistent, or when some women were paid more than some men. Subjects were particularly poor at recognizing discrimination when women's salaries were slightly less or equivalent to the men's salaries but the women were substantially better qualified along given salary-relevant dimensions.<sup>222</sup>

---

222. See Susan D. Clayton, The Recognition of Discrimination in a Minimal-Information Format, Paper Presented at the 97th Meeting of the American Psychological Association (August 1989),

Extending these results, Christel Rutte and her colleagues showed that people are quite poor at detecting discrimination when it is inconspicuous, or when information is presented in piecemeal rather than aggregate form.<sup>223</sup> In particular, their study showed that discrimination is especially difficult to detect when principles of "ordinal equity" are not violated.

Ordinal inequity exists when an employee who ranks higher on relevant input variables (e.g., qualifications) ranks lower than some other employee on the relevant outcome variables (e.g., salary, job title, rank or grade).<sup>224</sup> Magnitudes of difference do not implicate ordinal equity. In Rutte's study, when researchers gave advantages to majority group members in unobtrusive ways, not violating principles of ordinal equity, subjects generally failed to recognize patterns of disparate treatment. Perhaps more significantly, even where subjects recognized these less obtrusive patterns of sex-linked disparity, they were unlikely to characterize them as unfair.<sup>225</sup>

In summary, people expect discrimination to result in conspicuous, consistent patterns of outgroup subordination. But subtle forms of intergroup bias, while they will disadvantage the outgroup over time, are unlikely to produce such neat, consistent outcomes. We can reasonably predict that the kind of disadvantage resulting from cognitive forms of intergroup bias will frequently go unrecognized by fact finders (or perhaps even victims) as reflecting patterns of inequality. Even if fact finders recognize such patterns of disparity, absent violation of ordinal equity, they are unlikely to label them as unfair or discriminatory. Affirmative action programs, which emphasize the maintenance of numerical standards and the examination of broad patterns of distribution, have circumvented these problems. If affirmative action is eliminated, we cannot rely on remaining tools to achieve the same result.

#### *b. Schematic Expectancies and the Problem of Ingroup Favoritism*

When we think of discrimination, we think of discrimination *against* members of various outgroups. But as psychologist Marilyn Brewer observes, there are different types of intergroup bias.<sup>226</sup> From a psychological and policy standpoint, these differ in important ways.

---

described in *JUSTICE, GENDER, AND AFFIRMATIVE ACTION* 79-81 (Susan D. Clayton & Faye J. Crosby eds., 1992).

223. See Christel G. Rutte et al., *Organization of Information and the Detection of Gender Discrimination*, 5 *PSYCHOLOGICAL SCI.* 226 (1994).

224. See *id.* at 226.

225. See *id.* at 229-30.

226. See Marilyn B. Brewer, *In-Group Favoritism: The Subtle Side of Intergroup Discrimination*, in *CODES OF CONDUCT: BEHAVIORAL RESEARCH AND BUSINESS ETHICS* 160 (David M. Messick & Ann E. Tenbrunsel eds., 1996).

The first type Brewer identifies is bias or discrimination against the outgroup. In the case of Type I discrimination, the discriminator treats members of the outgroup unfairly, while members of the ingroup are treated neutrally with reference to some standard metric of fairness. Type II discrimination favors the ingroup. Here, treatment of the ingroup is biased in a positive or lenient direction, whereas treatment of the outgroup goes "by the book." Roberto da Matta encapsulates this type of discrimination in his description of Brazilian legal culture: "[f]or my friends, everything; for my enemies, the law!"<sup>227</sup> Type III involves both discrimination in favor of the ingroup and discrimination against the outgroup.

As noted earlier, both existing civil rights statutes and popular conceptions of discrimination assume Type I discrimination—negative treatment of outgroup members in relation to some objective metric of fairness. So, when a person claims to have been discriminated against, we tend to look for evidence that she was treated unfairly. For example, we look for deviations from established policies or procedures. If we find them, we interpret such deviations as evidencing discriminatory intent.<sup>228</sup> If the complainant does not appear to have been treated unfairly, we tend to conclude that no discrimination occurred.<sup>229</sup>

The tendency for discrimination disputes to center on the negative treatment of outgroup members, which no doubt reflects and reinforces the assumption of Type I discrimination, is driven both by the economics of discrimination litigation and by the effects of various cognitive biases. From an economic standpoint, unless a discrimination victim has suffered a substantial economic loss, such as employment termination, litigation will in most circumstances be unfeasible.

Psychological factors may play a role as well. People react more negatively to being deprived of something that they *have* had than to

---

227. See Roberto da Matta, *The Quest for Citizenship in a Relational Universe*, in *STATE AND SOCIETY IN BRAZIL* 307, 319 (John D. Wirth et al. eds., 1987) (reciting an ancient Brazilian expression).

228. See, e.g., *Lattimore v. Polaroid Corp.*, 99 F.3d 456, 466-67 (1st Cir. 1996) (holding plaintiff could show pretext by demonstrating that the defendant had deviated from established policies and practices); *Odima v. Westin Tuseon Hotel*, 53 F.3d 1484, 1492 (9th Cir. 1994) (holding as not clearly erroneous district court's finding of discrimination based in part on the fact that the defendant had deviated from established company policies and procedures); *Gaworski v. ITT Commercial Finance Corp.*, 17 F.3d 1104, 1119 (8th Cir. 1994) (holding defendant's failure to follow layoff and promotion policies is evidence from which jury could infer discrimination).

229. This implicit assumption of Type I discrimination may account in part for a phenomenon observed by sociologist Lauren Edelman and her colleagues in their study of corporate human resources professionals' approach to equal employment opportunity compliance. They observed that human resources professionals approach discrimination complaints as disputes about procedure, and tend to recast the obligations imposed by civil rights laws in procedural fairness terms. See Lauren B. Edelman et al., *Internal Dispute Resolution: The Transformation of Civil Rights in the Workplace*, 27 *LAW & SOC'Y REV.* 497, 512, 514 (1993).

being denied something that they have *never* had. Behavioral economists and cognitive psychologists refer to this phenomenon as the "endowment effect."<sup>230</sup> From this effect one can reasonably predict that people will more readily sue over a termination of employment than over a failure to hire or promote. Finally, as cognitive psychologists have demonstrated, people are generally less likely to notice things that *do not* happen than they are to notice things that *do* happen.<sup>231</sup> Being fired or subjected to other forms of employment discipline are far more salient than not being assigned a particular account, not being introduced to important people, or not receiving encouragement, training, or other forms of mentoring from an influential superior. Accordingly, the mobilization of civil rights remedies is much more likely to follow the occurrence of a negative event than the nonoccurrence of a positive event. This reinforces the underlying cultural assumption of Type I bias.

However, extensive empirical research in social cognition and social identity suggests that Type II bias largely underlies the kind of discrimination that results in the systematic advantaging of white males. This research divides into three broad types: research on racial attitudes, research on intergroup reward allocation, and research on intergroup helping behavior and the operation of leniency effects.

#### *i. Evidence of Type II Bias in Racial Attitudes Research*

The earliest empirical investigations of racial attitudes relied on subject self-report.<sup>232</sup> Recognizing that social desirability effects might distort findings in studies of this kind, researchers developed a new technique for measuring racial attitudes.<sup>233</sup> This technique, known as the response time study, is based on the previously verified insight that highly associated word pairs (e.g., chicken-egg), produce faster

---

230. See Daniel Kahneman & Amos Tversky, *Choices, Values, and Frames*, in JUDGMENT AND DECISION-MAKING 194, 207-08 (Hal R. Arkes & Kenneth R. Hammond eds., 1986) (discussing the cognitive psychology of loss aversion and resulting endowment effects); Richard Thaler, *Toward a Positive Theory of Consumer Choice*, 1 J. ECON. BEHAV. & ORG. 39, 44 (1980) (coining the term "endowment effect").

231. See, e.g., Lee Ross, *The Intuitive Psychologist and His Shortcomings: Distortions in the Attribution Process*, 10 ADVANCES IN EXPERIMENTAL SOCIAL PSYCHOLOGY 173, 196-97 (1977), reprinted in COGNITIVE THEORIES IN SOCIAL PSYCHOLOGY 337 (Leonard Berkowitz ed., 1978); Shelley E. Taylor & Susan T. Fiske, *Salience, Attention, and Attribution: Top of the Head Phenomena*, in 11 ADVANCES IN EXPERIMENTAL SOCIAL PSYCHOLOGY 249, 274 (Leonard Berkowitz ed., 1978).

232. See, e.g., Daniel Katz & Kenneth Braly, *Racial Stereotypes of One Hundred College Students*, 28 J. ABNORMAL & SOC. PSYCHOL. 280 (1933-34).

233. See, e.g., Harold Sigall & Richard Page, *Current Stereotypes: A Little Fading, A Little Faking*, 18 J. PERSONALITY & SOC. PSYCHOL. 247 (1971).

responses in word recognition tasks than word pairs that are not associated (e.g., doctor-butter).<sup>234</sup>

In one of the first response time studies, Samuel Gaertner and John McLaughlin paired the words "black" and "white" with various positive and negative trait constructs (e.g., smart, stupid, lazy, clean), and measured subjects' response times in a word recognition task.<sup>235</sup> Response times provided an unobtrusive measure of the strength of subjects' associations between the constructs "black" or "white" and various positive and negative traits. Fast response times indicated the presence of an association; slower response times indicated the absence of an association.

Gaertner and McLaughlin's subjects, who had previously measured "low prejudice" on a variety of scales, did not more readily associate negative trait concepts with the word "black." However, they reacted significantly more quickly when positive trait concepts were paired with the word "white."<sup>236</sup> These results suggest that although low prejudice individuals do not differentially associate negative traits with blacks, they do more readily associate positive traits with whites.

The results obtained in Gaertner and McLaughlin's early investigation are consistent with those of subsequent studies of the "ingroup priming effect." In these experiments, researchers showed that priming people with references to an outgroup does not systematically result in the activation of negative trait concepts. But priming them, even subliminally, with references to their ingroup does.<sup>237</sup> Dovidio, Evans, and Tyler found a similar effect using explicitly racial primes.<sup>238</sup> Rather than priming subjects with references to themselves or to a minimal ingroup, Dovidio and his colleagues primed subjects with references to the words "black" or "white." Priming subjects with references to "black" did not result in their responding more quickly to subsequently presented negative trait concepts. However, priming subjects with references to "white" did result in their responding more quickly to positive traits.

---

234. See David E. Meyer & Roger W. Schvaneveldt, *Facilitation in Recognizing Pairs of Words: Evidence of a Dependence Between Retrieval Operations*, 90 J. EXPERIMENTAL PSYCHOL. 227 (1971) (demonstrating relationship between association and response latency in word recognition tasks). For a thorough review of the history of racial attitude research, including the development of the response time study, see John F. Dovidio & Samuel L. Gaertner, *Prejudice, Discrimination, and Racism: Historical Trends and Contemporary Approaches*, in PREJUDICE, DISCRIMINATION & RACISM 1 (John F. Dovidio & Samuel L. Gaertner eds., 1986).

235. See Samuel L. Gaertner & John P. McLaughlin, *Racial Stereotypes: Associations and Ascriptions of Positive and Negative Characteristics*, 46 SOC. PSYCHOL. Q. 23, 24 (1983).

236. See *id.* at 27.

237. See, e.g., Crocker & Schwarz, *supra* note 123; C. W. Perdue et al., *Us and Them: Social Categorization and the Process of Intergroup Bias*, 59 J. PERSONALITY & SOC. PSYCHOL. 475 (1990). For a general discussion of these studies, see Bargh, *supra* note 111, at 9.

238. John F. Dovidio et al., *Racial Stereotypes: The Contents of Their Cognitive Representations*, 22 J. EXPERIMENTAL SOC. PSYCHOL. 22 (1986).

These results and others like them are consistent with the hypothesis that subtle forms of intergroup bias will manifest more as favoritism towards the ingroup than as hostility towards the outgroup.<sup>239</sup>

ii. *Type II Bias in Positive and Negative Distribution Decisions*

As described above,<sup>240</sup> much of our understanding of subtle forms of intergroup bias derives from the "minimal group" studies pioneered by social identity theorist Henri Tajfel and his colleagues.<sup>241</sup> Many of these studies explored patterns of intergroup preference in the allocation of monetary rewards. Researchers assigned subjects to groups on some trivial basis, then asked subjects who would not themselves benefit from allocation decisions to divide prize money between an ingroup and outgroup member.

In one of the earliest of these studies, researchers gave subjects a variety of allocation options, some of which favored the ingroup member, some of which favored the outgroup member, and some of which resulted in relatively equal allocations.<sup>242</sup> Subjects in this study, as in subsequent replications,<sup>243</sup> rejected those options that resulted in relatively equal allocations, and instead tended to choose those allocations which benefited the ingroup member at the expense of the outgroup member. As Marilynn Brewer notes however, one cannot tell from these results alone whether subjects are motivated to *favor* ingroup members or to *disfavor* outgroup members.<sup>244</sup> Thus, in and of themselves, the early ingroup reward allocation studies did not necessarily support the Type II bias hypothesis.

Subsequent studies investigating intergroup biases in the allocation of *negative* outcomes, however, do support the Type II bias hypothesis. In these studies researchers divided subjects into groups using the minimal group technique. They asked subjects to distribute negative outcomes, in the form of exposure to aversive physical stimuli, and positive outcomes, in the form of monetary rewards.<sup>245</sup> In the positive

---

239. For a thorough review of the racial and ingroup priming studies, see John F. Dovidio & Samuel L. Gaertner, *Stereotypes and Evaluative Intergroup Bias*, in *AFFECT, COGNITION, AND STEREOTYPING* 167 (Diane M. Mackie & David L. Hamilton eds., 1993).

240. See discussion in text accompanying note 155.

241. See generally Henri Tajfel, *Experiments in Intergroup Discrimination*, 223 *SCI. AM.* 96 (1970) (summarizing early work in this field).

242. See Tajfel et al., *supra* note 81.

243. For a thorough review of the intergroup reward allocation studies, including subsequent replications of Tajfel's original 1971 experiment, see Brewer, *supra* note 80, at 309-13.

244. See *id.*

245. See Amelie Mummendey et al., *Categorization is Not Enough: Intergroup Discrimination in Negative Outcome Allocation*, 28 *J. EXPERIMENTAL SOC. PSYCHOL.* 125 (1992); Miles Hewstone et al., *Social Categorization and Similarity in Intergroup Behaviour: A Replication with Penalties*, 11 *EUR. J. SOC. PSYCHOL.* 101 (1981).

outcome condition, subjects behaved as in previous studies: they chose the option that most benefited the ingroup member, even though the equal allocation option would have maximized the total amount of benefit distributed.<sup>246</sup>

In the negative allocation condition however, subjects behaved quite differently. When distributing *negative* outcomes, subjects preferred an equal allocation option that minimized the total amount of harm distributed. These patterns obtained even though subjects could have selected an option that modestly benefited the ingroup member and moderately harmed the outgroup member.<sup>247</sup> Evidently, it is one thing to benefit an ingroup member when it appears to leave the outgroup member no worse off, and quite another when ingroup gain can be achieved only at the obvious cost of outgroup harm. These findings too provide theoretical support for the plausibility of the Type II bias hypothesis. From them we can reasonably predict that employment decision makers are more likely to advantage a member of their own ingroup when they can do so without leaving an outgroup member worse off.

### *iii. Ingroup Helping Discrimination and the Leniency Effect*

Patterns of modern discrimination turn in large measure on the answer to one simple question: Who gets cut slack, and who does not? What happens when an employee violates a rule? Is she subjected to discipline under established policies, or are her transgressions overlooked, or attributed to factors beyond her control? When an ambiguous aspect of a person's background can be interpreted in various ways, one negative, one neutral, which attribution is made? And when a person simply needs help, does she receive it?

Type II bias in large measure shapes people's tendencies to assist or ignore, to excuse others' transgressions or hold them accountable under objective standards of conduct. The earliest, and perhaps still the most vivid of the studies illustrating this effect, was conducted in the

---

246. It is both interesting and important to note that subjects in these experiments rejected the Pareto optimal choice. (Borrowing the name of Italian economist Vilfredo Pareto, a distribution is said to be *Pareto optimal* or *Pareto efficient* if there is no way to change the distribution such that someone is made better off without making someone else worse off. Conversely, a distribution is *Pareto inefficient* if someone can be made better off without making anyone else worse off. See ROBERT COOTER & THOMAS ULEN, *LAW AND ECONOMICS* 12 (2d ed. 1997); A. MITCHELL POLINSKY, *AN INTRODUCTION TO LAW AND ECONOMICS* 7 n.4 (2d ed. 1989)). For an analysis of the implications of these and similar research findings for the law and economics approach to reflecting patterns of ingroup favoritism, see Richard H. McAdams, *Cooperation and Conflict: The Economics of Group Status Production and Race Discrimination*, 108 HARV. L. REV. 1003 (1995).

247. See Mummendey et al., *supra* note 245, at 133.



early 1970s, in front of a Kansas supermarket.<sup>248</sup> In this study, one black woman and one white woman, whom researchers matched for age and social class-related appearance, dropped a bag of groceries while leaving a supermarket, right in the path of oncoming shoppers. Researchers investigated whether white shoppers would help a white "bag dropper" more frequently than a black bag dropper.

The results were complex and intriguing. Overall, the experimenters found no significant effect of race on the provision of help *per se*: approximately the same percentage of incoming white shoppers in either condition stopped to help.<sup>249</sup> Subsequent analysis of the data however, revealed an important, if more subtle, phenomenon. When the bag dropper was white, sixty three percent of those who stopped continued to provide assistance until the job was done. When the bag dropper was black, helpers tended to pick up one or two items and then leave, providing complete assistance only thirty percent of the time.<sup>250</sup>

Additional studies provide further evidence of an ingroup helping bias. In 1977, Samuel Gaertner and John Dovidio conducted an experiment in which white subjects were led to believe that they were participating in an investigation of extrasensory perception (ESP).<sup>251</sup> Researchers assigned subjects to serve as either a "sender" or a "receiver," and paired each with a partner/confederate, who was either white or black. Senders and receivers sat in different rooms. Researchers told some subjects that a second person was sitting with their partner in the other room, and told other subjects that the partner was alone.

During the course of the "ESP" experiment, researchers staged an emergency. Subjects heard the sound of falling chairs and the screams of the partner in the other room, followed by prolonged silence. Researchers investigated whether the partner's race would effect the rates at which subjects would go to the aid of their partner.

When subjects believed that their partner was alone in the other room, the partner's race had no significant effect on responses.<sup>252</sup> However, when subjects believed that there was another person in the room with the partner, race made a dramatic difference. Where the apparently imperiled partner was white, seventy-five percent of subjects offered aid, but where the partner was black, the rate dropped to thirty-seven percent.<sup>253</sup> Perhaps even more significantly, subjects showed

---

248. See Lauren G. Wispe & Harold G. Freshley, *Race, Sex, and Sympathetic Helping Behavior: The Broken Bag Caper*, 17 J. PERSONALITY & SOC. PSYCHOL. 59 (1971).

249. See *id.* at 62.

250. See JAYNE ALLYN PILJAVIN ET AL., *EMERGENCY INTERVENTION* 158 (1981).

251. See Samuel L. Gaertner & John F. Dovidio, *The Subtlety of White Racism, Arousal, and Helping Behavior*, 35 J. PERSONALITY & SOC. PSYCHOL. 691 (1977).

252. See *id.* at 696.

253. See *id.*

greater physiological arousal, measured by change in heart rate, when the partner/confederate was white than when he was black.<sup>254</sup>

Gaertner and Dovidio interpreted these results as indicating that whites do not deliberately avoid providing assistance to blacks. However, when features of the situation are ambiguous, when it is unclear whether help is called for, they tend to resolve uncertainty in favor of helping whites and against helping blacks.<sup>255</sup>

Similar results occur when the ambiguity concerns whether help is *deserved* as opposed to whether help is *needed*. So, where the need for help on a particular task is attributable to a lack of effort, white subjects provide help to white confederates at a significantly higher rate than they provide to similarly situated black confederates.<sup>256</sup>

As the tendency to assist is biased, so is the tendency to overlook or excuse transgression. For example, in a 1974 field study of whites' reactions to apparent shoplifting, Max Dertke and his colleagues demonstrated that when the shoplifter/confederate was black, white shoppers spontaneously reported and followed up on an observed shoplifting incident at a much higher rate than when the shoplifter/confederate was white.<sup>257</sup>

A more recent study illustrates a similar Type II leniency effect in the attribution of positive and negative behaviors.<sup>258</sup> Using the minimal group technique, Joseph Weber divided subjects into groups and then had them watch ingroup or outgroup confederates either perform or fail to perform an altruistic act. Where the confederate performed the altruistic act, researchers observed no intergroup bias; subjects even-handedly attributed the confederate's action to positive personality traits, such as helpfulness or generosity. Intergroup bias surfaced only when the confederate had *failed* to act. When an outgroup member failed to act, subjects attributed the failure to negative personality traits. But when a member of the ingroup failed to act, subjects explained the failure as resulting from situational constraints.<sup>259</sup>

One can easily see how over time, these subtle forms of ingroup favoritism would result in markedly different outcomes for ingroup and outgroup members. If decision makers react to members of their own social reference groups with more positive associations, a quicker

---

254. See *id.* at 698.

255. See *id.* at 699.

256. See Samuel L. Gaertner & John F. Dovidio, *The Aversive Form of Racism*, in PREJUDICE, DISCRIMINATION AND RACISM 61, 70 (John F. Dovidio & Samuel L. Gaertner eds., 1986).

257. See Max C. Dertke et al., *Observer's Reporting of Shoplifting as a Function of Thief's Race and Sex*, 94 J. SOC. PSYCHOL. 213 (1974).

258. See Joseph G. Weber, *The Nature of Ethnocentric Attribution Bias: Ingroup Protection or Enhancement?*, 30 J. EXPERIMENTAL SOC. PSYCHOL. 482 (1994).

259. See *id.* at 491.

willingness to help, and a stronger inclination to ignore or excuse shortcomings, it is easy to predict who will be systematically advantaged in hiring and promotion decisions. Disparities will develop even absent hostile animus or negative actions directed towards the outgroup.

*c. In-Group Favoritism and the Limitations of Current Law*

As it is increasingly being interpreted, Title VII is poorly equipped to control discrimination resulting from ingroup favoritism. Consider for example the problem of word-of-mouth recruitment, a process by which current employees disseminate to their friends information about job opportunities or suggest those friends to hiring officials as a means of identifying candidates for openings. Early Title VII cases almost uniformly found such hiring procedures discriminatory on the grounds that employees tend to recommend individuals of their own race and ethnicity, thereby perpetuating patterns of racial or ethnic imbalance.<sup>260</sup> Over time however, this consensus has eroded. There now exists a split in the circuits as to whether word-of-mouth recruitment violates Title VII, even under a disparate impact analysis.<sup>261</sup>

The cases refusing to impose liability for word-of-mouth recruitment vividly illustrate the courts' increasing preoccupation with Type I discrimination and their corresponding inability to recognize the discriminatory effect of Type II bias. For example, in *E.E.O.C. v. Chicago Miniature Lamp Works*, Judge Cummings opined:

---

260. See, e.g., *Barnett v. W.T. Grant Co.*, 518 F.2d 543, 549 (4th Cir. 1975) (finding word-of-mouth hiring discriminatory because of its tendency to perpetuate the all-white composition of a work force); *Franks v. Bowman Transportation Co., Inc.* 495 F.2d 398, 419-420 (5th Cir. 1974), *rev'd on other grounds*, 424 U.S. 747 (1976) (holding word-of-mouth recruitment discriminatory on a disparate impact theory); *Rowe v. General Motors Corp.*, 457 F.2d 348, 359 (5th Cir. 1972) (stating white employees tend not to know African Americans who would be qualified to perform the work in question); *Parham v. Southwestern Bell Tel. Co.*, 433 F.2d 421, 427 (8th Cir. 1970) (finding existing white employees tend to refer members of their own race, thus perpetuating racial disparities); *NAACP v. City of Corinth*, 83 F.R.D. 46, 62 (N.D. Miss. 1979) (holding prima facie case of disparate treatment discrimination established by showing that job vacancies were filled by word-of-mouth).

261. Compare *Domingo v. New England Fish Co.*, 727 F.2d 1429, 1435-36 (9th Cir. 1984) (holding word-of-mouth recruitment in white-dominated workforce discriminatory), and *NAACP v. Evergreen*, 693 F.2d 1367, 1369 (11th Cir. 1982) ("[W]ord of mouth [recruitment] undoubtedly operated to the benefit of white applicants and to reduce the number of potential black applicants. . ."), with *Equal Employment Opportunity Comm'n v. Consol. Service Sys.*, 989 F.2d 233, 235-36 (7th Cir. 1993) (holding word-of-mouth recruitment does not violate Title VII on either a disparate treatment or disparate impact theory; it was the most cost-effective form of recruitment, and there was no evidence of invidious bias against any under-represented group), *Equal Employment Opportunity Comm'n v. Chicago Miniature Lamp Works*, 947 F.2d 292, 298-99 (7th Cir. 1991) (holding word-of-mouth recruitment does not violate Title VII), and *Equal Employment Opportunity Comm'n v. O&G Spring & Wire Forms Specialty Co.*, 732 F. Supp. 72, 73-74 (N.D. Ill. 1990) (holding word-of-mouth recruitment is cost free and produces a high percentage of skilled applicants).

Miniature's passive reliance on word-of-mouth to generate applicants must be given minimal weight because it involved no affirmative act by Miniature. Drawing the inference of intent from "non-action" is inherently more difficult than drawing the inference of intent from particular actions. This is especially true since intent means more than knowledge that a certain action (or non-action) will cause certain discriminatory results. Intent means a subjective desire or wish for these discriminatory results to occur.<sup>262</sup>

Similarly oblivious to the significance of subtle forms of ingroup preference involved in "non-action" towards minorities, Judge Posner writes in *E.E.O.C. v. Consolidated Service Systems* that if word-of-mouth recruitment "just happens to produce a work force whose racial or religious or ethnic or nation-origin or gender composition pleases the employer, this is not intentional discrimination. The motive is not a discriminatory one."<sup>263</sup> It would be difficult to find more vivid illustrations of the assumption of Type I bias reflected in Title VII case law.

Thus, even if a nonrecruited individual recognizes that word-of-mouth recruitment has deprived her of a potential employment opportunity, and even if such practices serve to perpetuate existing racial or other imbalances in an employer's workforce, an increasing number of courts refuse to provide a Title VII remedy. Judicial interpretations of Title VII are increasingly based on the application of prototypical expectancies rather than on clear thinking about what intergroup bias is, what causes it, and how it influences social judgment. Until our civil rights jurisprudence is premised upon a more complete and accurate understanding of intergroup bias, those forms of discrimination which differ from the most superficial schematic expectancies will continue to go unrecognized and unaddressed.

Ingroup favoritism manifests itself gradually in subtle ways. It is unlikely to trigger mobilization of civil rights remedies because instances of this form of discrimination tend to go unnoticed. If they are noticed, they will frequently seem genuinely trivial or be economically unfeasible to pursue. By the time some more salient event, such as a layoff decision or a failure to promote, precipitates a civil rights challenge, the preferred ingroup and nonpreferred outgroup members may no longer be similarly situated. Having profited over time from a series of subtle, incremental advantages, the ingroup member is apt to be objectively better situated. For this reason as for others, we cannot expect existing equal opportunity tools adequately to prevent, identify, or redress this more modern form of discrimination. Eliminating

---

262. 947 F.2d at 298.

263. 989 F.2d at 235.

affirmative action without first repairing these deficiencies is a recipe for racial, ethnic, and gender injustice.

*D. Individualized Adjudication, Hypothesis Testing, and Causal Attribution: The Effects of Intergroup Bias*

Consider for a moment the judgment task involved in adjudicating an individual employment discrimination suit. Determining in any given case whether discrimination has occurred is fundamentally an exercise in causal attribution. The employer has taken some negative action, most frequently a termination of employment, against the plaintiff. The jury's role is to determine why that negative action was taken. Was it, as the plaintiff alleges, because the decision maker discriminated against her because of her membership in a protected group? Or was it, as the defendant argues, because of some legitimate, nondiscriminatory reason, usually some malfeasance or deficiency on the plaintiff's part? In a hiring case, did the plaintiff fail to get the job because the decision maker took her group status into account in making the challenged decision? Or did the decision maker believe that some other candidate would do a better job? The trial of such a case will essentially entail a battle between two competing causal theories. Seeking to convince the jury that discrimination is to blame, the plaintiff will portray the decision makers as discriminators. Seeking to convince the jury that the plaintiff is to blame, the defendant will do everything possible to make his or her deficiencies salient.

Why should we expect a jury to approach this social decision task free from the various forms of intergroup bias that distort intergroup perception and judgment in other contexts? As we have seen, unconscious stereotypes about members of different social groups create implicit expectancies in the minds of social perceivers. These expectancies in turn distort the perception, interpretation, and recall of information about members of the targeted groups, pulling subsequent social judgments in a stereotype-consistent direction. Stereotypic expectancies and other forms of intergroup bias also affect causal attribution, causing unconscious distortions in the interpretation and perceived predictiveness of past behavior.<sup>264</sup>

Attributing the causes of employment decisions implicates the very processes of social perception and judgment bound up in the challenged employment decisions themselves. Unless the demographic characteristics of fact finders vary in some dramatic way from those of the

---

264. Attributions to stable internal factors increase the predictive value subjectively assigned to an event. Attributions to transient external factors tend to diminish perceived predictiveness. See Bernard Weiner et al., *Perceiving the Causes of Success and Failure*, in *Attribution* 95, 113-17 (Edward E. Jones et al. eds., 1971).

decision makers, we cannot reasonably expect that the level of intergroup discrimination reflected in employment decisions will vary in any meaningful way from the level reflected in discrimination verdicts. Indeed, the analytical structure and content of disparate treatment adjudications, focusing as it does on the plausibility of defendant's proffered legitimate nondiscriminatory reasons for a challenged employment decision, can be expected to potentiate those forms of intergroup bias caused by stereotypic expectancies.<sup>265</sup> Discrimination adjudications therefore may be even more vulnerable to cognitive forms of intergroup bias than the decision tasks which give rise to them.

An employer's determination, for example, whether a particular employee should be terminated is not much different from a court's determination whether an employer believed in good faith that a particular employee deserved to be terminated. Similarly, an employer judging whether a particular candidate is best qualified for a position is not particularly different from a court judging whether a particular candidate would reasonably have been viewed by a well-intentioned employer as the best qualified person for a position. Expectancy confirmation effects, such as those illustrated in Darley's and Gross's study discussed above,<sup>266</sup> will distort both types of judgments. Thus, there is no reason to believe that the incidence of stereotype-induced judgment error in discrimination adjudications will differ in any significant way from its incidence in employment or educational decision making. Indeed, if as Darley and Gross suggest, exposure to ambiguous but ostensibly diagnostic collections of information potentiates expectancy confirmation bias, we can expect disparate treatment adjudications, with their "information rich texture," to suffer even more from such biases than hiring or educational admissions decisions, where relatively little diagnostic information is available and decision makers may be more on guard against making stereotypic judgments.

In short, from a cognitive process standpoint we cannot expect disparate treatment adjudications to be any less subject to subtle forms of intergroup bias than the decisions which give rise to them. Correspondingly, we cannot expect individualized adjudication of disparate treatment claims to be particularly effective in identifying or redressing cognitive discrimination. For this reason as for others, disparate treatment adjudication, like the colorblindness model of nondiscrimination

---

265. The nature and allocation of the parties' burdens of proof in Title VII individual treatment cases are set out in a trilogy of well-known Supreme Court decisions, and are so frequently recited that I will not reiterate them here. For those unfamiliar with the relevant doctrine, and with the significance of the employer's proffer, see *St. Mary's Honor Center v. Hicks*, 509 U.S. 502 (1993); *Texas Dept. of Community Affairs v. Burdine*, 450 U.S. 248, 254 (1981); *McDonnell Douglas Corp. v. Green*, 411 U.S. 792, 802-04 (1973).

266. See discussion in text accompanying notes 220-221.

and reliance on an objective concept of merit, is an extremely weak tool for combating cognitive forms of intergroup bias. We cannot expect these policies to do the work once accomplished by disparate impact theory, numerical standards, and the systematic, self-critical analysis of selection procedures. Unfortunately, when a person is color-blind, there is simply much he will not see.

#### CONCLUSION:

#### SOCIAL COGNITION, SOCIAL IDENTITY, AND THE ECOLOGY OF INTERGROUP BIAS

At its core, our antidiscrimination law and policy suffers from a cognitive bias known to psychologists as the fundamental attribution error.<sup>267</sup> This interpretive tendency, so pervasive in Anglo-American culture as to be called fundamental, leads people to over-attribute social behavior to stable, dispositional traits and to underestimate the causal efficacy of situational determinants of judgment, behavior, and choice.

Current antidiscrimination theories, most particularly disparate treatment theory, locate the problem of discrimination within the individual decision maker and largely overlook the role of the social environment in which those decision makers perceive, judge, and choose. In other words, when we ask, "Why does discrimination occur?" we tend to focus on stable, dispositional factors internal to actors we call "discriminators," rather than on characteristics of the situation in which decision makers function.

This is a serious mistake. Although we certainly need to know more, we know a good deal about the conditions under which intergroup bias will tend to increase or subside. We know, for example, that when people are asked to make judgments about the characteristics associated with different classes of people, they are less prone to illusory correlation—the drawing of unfounded, stereotype-consistent associations—when they actually view, arrange, and make notes about the data from which judgments are to be made.<sup>268</sup>

Similarly, experimental evidence suggests that we can reduce stereotypic biases in evaluation when evaluative criteria are specifically identified and when we provide decision makers with a rich body of specific diagnostic information on which to base decisions.<sup>269</sup> Stereotypic expectancies and other subtle forms of bias will have less influence in

---

267. See Lee Ross, *The Intuitive Psychologist and His Shortcomings: Distortions in the Attribution Process*, 10 *ADVANCES IN EXPERIMENTAL SOCIAL PSYCHOLOGY* 173 (Leonard Berkowitz ed., 1977) (coining the term "fundamental attribution error").

268. See Loren J. Chapman & Jean P. Chapman, *Genesis of Popular but Erroneous Psychodiagnostic Observations*, 72 *J. ABNORMAL PSYCHOL.* 193, 202-04 (1967).

269. See generally Veronica F. Nieva & Barbara A. Gutek, *Sex Effects on Evaluation*, 5 *ACAD. MGMT. REV.* 267, 271 (1980) (reviewing relevant research).

the evaluation process when decision makers apply less generalized, and more specific, preferably objective, criteria.<sup>270</sup> Requiring decision makers to "consider the opposite," that is, to make the case for the option they are *not* selecting may reduce the biasing effect of prior theories, including the biasing effect of social stereotypes functioning as implicit schematic expectancies.<sup>271</sup> Similarly, requiring evaluators to summarize the evidence tending to contradict their conclusions may weaken somewhat the subsequent biasing effect of the implicit theory created by the evaluation itself.<sup>272</sup>

We know that whites tend to discriminate less against non-whites in the allocation of negative sanctions if they know that peers will monitor and potentially censure their behavior.<sup>273</sup> Monitoring of this sort is most effective when individual evaluators are held individually responsible for outcomes. The diffusion of responsibility associated with group decision making appears to lessen the impact of review and potential censure.<sup>274</sup>

We know that these situational factors affect the likelihood that cognitive forms of discrimination will occur. But none of this knowledge is incorporated into existing theories of discrimination or into the models of proof we provide to those fact finders charged with identifying it. Our jurisprudence relentlessly directs these fact finders towards "discriminatory intent," a stable, trait-like characteristic located somewhere inside decision makers' minds, functionally detached from the perceptual environment in which those decision makers perceive, judge, and choose.

On a larger scale, we also know something about how the composition and social organization of a particular institution or community will influence intergroup perception, judgment, and reward allocation. For example, the presence of a "token" woman or minority does little to lessen patterns of intergroup bias in an otherwise homogeneous group. In fact, a "solo" in such a situation operates at a much elevated risk of discrimination.<sup>275</sup>

---

270. On the value of decomposition as a means of reducing evaluation bias, see, e.g., J. Scott Armstrong et al., *The Use of the Decomposition Principle in Making Judgments*, 14 *ORG. BEHAV. HUM. PERFORMANCE* 257 (1975).

271. See, e.g., Craig A. Anderson, *Inoculation and Counterexplanation: Debiasing Techniques in the Perseverance of Social Theories*, 1 *SOC. COGNITION* 126 (1982) (exploring the consideration of alternatives as a technique for limiting the biasing effect of causal theories).

272. See *id.*

273. See Edward Donnerstein & Marcia Donnerstein, *Variables in Interracial Aggression: Potential Ingroup Censure*, 27 *J. PERSONALITY & SOC. PSYCHOL.* 143, 148-50 (1973).

274. See Samuel L. Gaertner & John F. Dovidio, *Racism Among the Well-Intended*, in *PLURALISM, RACISM, AND PUBLIC POLICY* 208, 217-18 (Edwin G. Clausen & Jack Bermingham eds., 1981).

275. On the operation of "solo effects" in social evaluation, see, e.g., David L. Hamilton & Tina K. Trolier, *Stereotypes and Stereotyping: An Overview of the Cognitive Approach*, in *PREJUDICE*,



We also know that simple integration does little to reduce intergroup conflict. In fact, where intergroup competition is present, contact between groups exacerbates, rather than reduces, intergroup bias.<sup>276</sup> At a minimum, for intergroup conflict to subside, members of different groups must become involved in relationships of cooperative interdependence.<sup>277</sup> Some researchers argue that even this is not enough: so long as category boundaries remain salient, even a cooperative intergroup reward structure will fail to reduce intergroup bias.<sup>278</sup> According to this view, the only way to reduce intergroup discrimination is to lessen the salience of intergroup distinctions. So long as social category boundaries remain salient, intergroup discrimination will persist.

And so we are led to the crucial question: how as a society can we reduce the salience of intergroup distinctions based on race, sex, national origin, or religion? One thing by now I hope is clear: we cannot accomplish this objective by simply directing people not to notice or take these distinctions into account. We cannot accomplish it by criminalizing discrimination, as some in Congress now seek to do.<sup>279</sup> We cannot accomplish it through even the vigorous enforcement of existing civil rights protections. They are too weak; they too fundamentally misapprehend the etiology and natural history of intergroup bias. And finally, we will not accomplish this objective either by eliminating or by preserving preferential forms of affirmative action.

For better or for worse, the application of insights from social cognition and social identity theory complicates rather than simplifies the affirmative action debate. On the one hand, there is reason to fear that preferential forms of affirmative action, at least in some contexts,

---

DISCRIMINATION & RACISM, 127, 134-36 (John F. Dovidio & Samuel L. Gaertner eds., 1986); Shelley E. Taylor, *A Categorization Approach to Stereotyping*, in *COGNITIVE PROCESSES IN STEREOTYPING AND INTERGROUP BEHAVIOR* 83, 90-94 (David L. Hamilton ed., 1981).

276. See, e.g., Bass & Duntzman, *supra* note 87; Marilynn B. Brewer & Roderick M. Kramer, *The Psychology of Intergroup Attitudes and Behavior*, 36 ANN. REV. PSYCHOL. 219, 227 (1985). Tajfel assigned particular emphasis to the role of competition in exacerbating intergroup conflict. See Henri Tajfel, *Social Categorization, Social Identity, & Social Comparison*, in *DIFFERENTIATION BETWEEN SOCIAL GROUPS* 61 (Henri Tajfel ed., 1978).

277. For research finding a positive effect from cooperative interdependence, see Stuart W. Cook, *Cooperative Interaction in Multiethnic Contexts*, in *GROUPS IN CONTACT* 155 (Norman Miller & Marilynn B. Brewer eds., 1984); David W. Johnson, et al., *Interpersonal Attraction Among Heterogeneous and Homogeneous Individuals: A Theoretical Foundation and a Meta-Analysis of the Research*, 53 REV. EDUC. RES. 5 (1983).

278. See Marilynn B. Brewer & Norman Miller, *Beyond the Contact Hypothesis: Theoretical Perspectives on Desegregation*, in *GROUPS IN CONTACT* 281 (Norman Miller & Marilynn B. Brewer eds., 1984); Norman Miller & Marilynn B. Brewer, *Categorization Effects on Ingroup and Outgroup Perception*, in *PREJUDICE, DISCRIMINATION, AND RACISM* 209, 227-28 (John F. Dovidio & Samuel L. Gaertner eds., 1986).

279. See, e.g., The Anti-Discrimination Act of 1977, H.R. 2080, 105th Cong. (1997) (proposing amendment to Section 706 of the Civil Rights Act of 1964 that would impose criminal liability for discrimination under Sections 703 or 704).

may indeed exacerbate intergroup tensions and perpetuate rather than reduce subtle forms of intergroup bias. In other words, preferences appear to do harm as well as good.

On the other hand, insights derived from these fields suggest that we are not yet ready to abandon preferential forms of affirmative action for the simple reason that we have nothing adequate with which to replace them. Neither our political discourse nor our civil rights jurisprudence reflect an adequate understanding of what intergroup discrimination is, how or what causes it to occur, or what is required to reduce or eliminate it. We underestimate the stability, subtlety, and perniciousness of intergroup bias. Perhaps most importantly, we lack a coherent theory as to why improving intergroup relations should be a high priority in American political, economic, and cultural life.

These failures, and the confidence of anti-affirmative action activists, derive at least in part from two widely held but ultimately erroneous assumptions. The first assumption is that discrimination is conscious, intentional, and reasonably easy to identify. According to this view, by monitoring their thought processes, well-meaning, law-abiding people can and will refrain from discriminating. When discrimination does occur, it can be readily identified and negatively sanctioned. Nothing more than vigorous enforcement of existing laws prohibiting discrimination is needed to pave the way to an equal opportunity society.

The second assumption holds that absent state-sanctioned or overt discrimination by private actors (which most would agree are waning), intergroup relations in the United States will improve or at least remain relatively tranquil, of their own accord. Intergroup harmony, rather than intergroup strife, is seen as a kind of default setting, necessitating little political intervention beyond the enactment and vigorous enforcement of laws prohibiting overt discrimination.

But tendencies towards intergroup discrimination are much more subtle, stable, and pernicious than these assumptions or their reflection in anti-affirmative action rhetoric admit. Social cognition teaches that much intergroup discrimination is both unintentional and unconscious. It occurs spontaneously as an unwanted artifact of normal cognitive functions associated with the processing of information about other people and can be corrected, if at all, only through further deliberate mental effort. Social identity theory and related research in experimental social psychology indicates that the tendencies to assist or excuse those with whom we feel closely identified and to subordinate the socially distant are far less tractable than we might wish.

These more subtle, incremental forms of discrimination are difficult to recognize, and neither our cultural understanding nor our

jurisprudential models of discrimination illuminate or provide ways to reckon with them. Without affirmative action, it remains to be seen how powerfully they will operate to exclude minorities and women from large segments of the academy, public contracting, or labor markets. But there is ample reason to fear that the road ahead may be rockier than Proposition 209's supporters optimistically predict.

One need only consider recent events in Bosnia, recall the burning cities of the nineteen-sixties, or reflect upon the great humanitarian disasters scarring human history to recognize that intergroup relations stand poised in delicate balance. Tensions can flare suddenly, violently, and with tragic consequences. We may agree that preferential forms of affirmative action come at a cost. We may aspire to replace them with something more closely reflecting our highest values and ideals. But to abandon them now, without an adequate theory of discrimination or workable equal opportunity policy to replace them, seems a risky course indeed.